nature communications

Article

https://doi.org/10.1038/s41467-025-67217-2

Intrinsically disordered regions facilitate target search to drive promoter selectivity by a yeast transcription factor

Received: 14 July 2025

Accepted: 24 November 2025

Published online: 10 December 2025

Check for updates

Nir Strugo[1,2], Carmit Burstein[1], Sk Saddam Hossain[3], Noam Nago[1], Hadeel Khamis[1] & Ariel Kaplan[1,2,4]

Transcription factors regulate gene expression by binding specific DNA motifs, yet only a fraction of putative sites is occupied in vivo. Intrinsically disordered regions have emerged as key contributors to promoter selectivity, but the underlying mechanisms remain incompletely understood. Here, we use single-molecule optical tweezers to dissect how disordered regions influence DNA binding by Msn2, a yeast stress-response regulator. We show that these regions power a search mechanism, facilitating initial non-specific association with DNA and promoting one-dimensional scanning toward target motifs, supported by charge-mediated interactions. Remarkably, this mechanism displays sequence sensitivity, with promoter-derived sequences enhancing both initial binding and scanning rates, demonstrating that Msn2–DNA interactions alone are sufficient to confer promoter selectivity in the absence of chromatin or cofactors. Our findings provide direct mechanistic evidence for how intrinsically disordered regions tune transcription factor search dynamics for Msn2 and expand sequence recognition beyond canonical motifs, supporting promoter selectivity in complex genomic contexts.

Eukaryotic transcription factors (TFs) orchestrate gene expression by binding to specific DNA motifs in promoters and enhancers, yet only a subset of these short and degenerate motifs present in the genome is occupied in vivo[1,2]. Recent work has highlighted a critical role for intrinsically disordered regions (IDRs) in conferring promoter specificity, showing that many yeast TFs, including zinc-finger, bZIP, and homeodomain proteins, require extensive IDRs for selective promoter binding[3–5]. This IDR-mediated selectivity implies an additional layer of specificity beyond the canonical DNA-binding domain (DBD), but the molecular mechanisms underlying the IDRs' role remain elusive. Characterized by their lack of secondary or tertiary structure, IDRs exhibit high conformational flexibility and multivalency, which allow them to participate in dynamic and context-dependent interactions. IDRs have been implicated in diverse regulatory functions in vivo: they drive transcriptional activation through recruitment of coactivators, assemble phase-separated condensates that concentrate transcriptional machinery, and modulate chromatin residence times[6,7]. In mammalian cells, disordered regions of TFs such as p53, NFκB, and the glucocorticoid receptor contribute to spatial confinement or prolonged chromatin engagement[8–10]. However, the direct contributions of IDRs to TF–DNA interactions and promoter selectivity remain untested.

TF–DNA binding is governed not only by binding affinity, but also by the kinetics of how sites are located and engaged[11]. According to the facilitated diffusion model, TFs alternate between three-dimensional diffusion in solution and one-dimensional exploration along the DNA, which can occur through continuous sliding, short-range hopping, or longer-range intersegmental transfer and looping events. By combining these modes, TFs reduce the dimensionality of their search and enhance the rate of motif recognition[12,13]. As they scan DNA, TFs must

[1]Faculty of Biology, Technion – Israel Institute of Technology, Haifa, Israel. [2]Russell Berrie Nanotechnology Institute, Technion – Israel Institute of Technology, Haifa, Israel. [3]Department of Chemical and Structural Biology, Weizmann Institute of Science, Rehovot, Israel. [4]Faculty of Biomedical Engineering, Technion – Israel Institute of Technology, Haifa, Israel. e-mail: akaplanz@technion.ac.il

also discriminate between cognate and non-cognate sequences, a process that has been shown to involve kinetic features such as transient contacts, dwell-time heterogeneity, and multistate binding behaviors[14–17]. Recent single-molecule studies have further revealed that sequence specificity can arise primarily from differences in association rates rather than dissociation kinetics, and that TFs can explore broad DNA surfaces and bypass potential targets during search[16,18,19]. IDRs may modulate these kinetic processes by promoting initial electrostatic engagement, stabilizing transient complexes, or tuning one-dimensional diffusion dynamics[20,21]. In parallel, several studies have shown that TF binding in vivo is influenced by DNA sequence features outside the core motif, such as flanking sequence composition, DNA shape, or chromatin architecture[22–24]. These observations raise the possibility that disordered regions enable TFs to sense contextual features and bias their search toward biologically relevant sites. Yet, a precise, quantitative dissection of IDR contributions to association rates, residence times, and sequence-dependent target-search behavior at the single-molecule level remains incomplete.

To directly address these questions, we use single-molecule optical tweezers to investigate how IDRs influence DNA binding by Msn2, a master regulator of the environmental stress response in budding yeast[25]. *S. cerevisiae* Msn2, composed of 704 amino acids, contains a 62-residue canonical DBD that targets stress-responsive promoter elements, flanked by extensive IDRs[26] (Fig. 1a) required for its in vivo promoter specificity and regulatory function[3]. Building on our previous single-molecule studies of transcription factor binding dynamics[27,28], we developed assays to probe specific interactions (binding at the cognate recognition motif) and non-specific interactions (association with other DNA regions), as well as Msn2's ability to engage in one-dimensional (1D) scanning along DNA. These experiments allow us to dissect how disordered regions contribute to promoter-selective binding and to determine their specific roles in TF–DNA recognition.

Results

IDRs increase Msn2 binding affinity via charge-mediated interactions

To identify the contribution of IDRs to DNA binding, we compared the binding performance of wild-type Msn2 (WT) and an IDR-truncated mutant (DBD) using DNA unzipping with optical tweezers (Fig. 1b; see also Supplementary Fig. 1 and Supplementary Table 1). In previous work[27], we showed that a bound TF hinders the propagation of the DNA unzipping fork, allowing us to measure the protein's position and breaking force. Repeated measurements, where the system is allowed to thermally equilibrate between iterations, reveal the binding probability, a direct measure of the TF's affinity for its binding site. In initial experiments, we used a section of the *Hor7* gene promoter, which contains four AGGGG binding motifs (Supplementary Fig. 2a) and is known to be bound by Msn2 in vivo[3]. However, the relatively small force signature of Msn2, which possesses only two zinc-fingers, combined with the high-force background from the flanking sequences, hindered the identification of binding events, and we were able to observe events only at one of the four sites (Supplementary Fig. 2a). To improve sensitivity, we designed a dedicated DNA construct with a single Msn2 binding motif, flanked by a short AT-rich sequence (which requires lower unzipping forces) and preceded by a ~300 bp segment lacking binding motifs. This design minimized background forces and enabled clearer detection of binding events (Fig. 1c). Repeated unzipping experiments with WT and DBD revealed a significant decrease in binding probability upon IDR removal (Fig. 1d, top), indicating that Msn2's IDRs enhance its binding affinity for the binding motif ($\Delta\Delta G \sim 2\,k_B$T; Supplementary Table 18). The breaking force for the bound complex was reduced too (Fig. 1d, bottom). Notably, we also tested a variant containing only the IDRs (IDR; Supplementary Fig. 1 and Supplementary Table 1) and observed no detectable unzipping signature at the motif.

Next, we sought to understand the mechanism by which IDRs increase the affinity of Msn2 for DNA. We hypothesized that charge-mediated interactions might contribute and initially tested this by increasing the ionic strength of the buffer with additional KCl (from 150 mM to 200 mM). This treatment reduced the binding probability and breaking force for WT Msn2 (Supplementary Fig. 2b), consistent with electrostatic screening. To further probe this mechanism, we introduced 50 mM free L-arginine into the buffer (which also contains 150 mM KCl). At pH 8.0, L-arginine is positively charged due to protonation of its guanidino and amino groups, while its carboxyl group is deprotonated (Fig. 1e and Supplementary Tables 2, 3). This treatment also led to a marked reduction in binding probability for WT Msn2 (Fig. 1f, top; Supplementary Fig. 2b, c). However, since L-arginine can also engage directly with DNA or protein surfaces, alter solvation and affect IDR conformational preferences, we further tested whether the effect depends specifically on its net charge. Raising the pH to 9.8, which neutralizes the amino group of L-arginine (pKa = 9.0) without substantially affecting the side chains in Msn2 (Supplementary Table 3), partially restored binding. At this higher pH, the binding probability was similar regardless of whether L-arginine was present, supporting the conclusion that the inhibition observed at pH 8.0 reflects electrostatic screening by the charged form of L-arginine, rather than other arginine-induced effects. Nonetheless, binding was reduced overall at pH 9.8 compared with pH 8.0, even in the absence of L-arginine, an effect we attribute to partial deprotonation of side chains within the IDR, possibly enhanced by local shifts in pKa due to charge regulation. A more detailed analysis of the effects of L-arginine and pH is provided in the Supplementary Discussion. The DBD variant showed a qualitatively similar trend (Fig. 1f, bottom; Supplementary Fig. 2b, c), consistent with the known electrostatic interactions between zinc fingers and DNA, but the effect was substantially milder. Hence, our results suggest that the IDRs enhance DNA binding largely through charge-mediated interactions. Moreover, they establish L-arginine as a useful experimental probe for selectively perturbing IDR–DNA interactions, enabling further dissection of the role of disordered regions in TFs function.

IDRs enhance the association rate of Msn2 to its binding motif

While our unzipping experiments demonstrated that IDRs increase Msn2's overall binding affinity, they did not reveal whether this enhancement reflects changes in association rate (k_{on}), dissociation rate (k_{off}), or both. To dissect the specific kinetic parameters affected by IDRs, we exploited a method we previously developed[28], based on monitoring thermal breathing fluctuations of DNA (Fig. 2a). The DNA construct is unzipped until reaching several base-pairs upstream of the binding motif, after which the optical traps are held at a constant position. The DNA then undergoes rapid thermal fluctuations on the millisecond timescale, during which approximately 20 bp, including the binding site, transiently and repeatedly separate into single strands ("open") and re-anneal ("close"). TF binding traps the DNA in the closed state, thereby suppressing these fluctuations and allowing for the identification of the bound complex. By analyzing the durations of unbound and bound intervals, we extract the association and dissociation rates, respectively.

Using this approach with WT and DBD (Fig. 2b), we found that while the dissociation rate was unaffected by IDR truncation, the association rate decreased approximately sixfold (Fig. 2c), indicating that IDRs enhance binding affinity by accelerating target association (resulting in $\Delta\Delta G \sim 1.6\,k_B$T, consistent with the equilibrium measurements; Supplementary Table 18). Notably, this kinetic signature is inconsistent with a strictly two-state (free ↔ bound) model and instead suggests the presence of at least one intermediate along the binding pathway, such as a transient encounter or partially engaged state. In this view, the IDRs facilitate the initial formation of such intermediates, increasing the overall association rate, while the stability of the final

bound complex remains determined by the DBD. Notably, the identical zero-force dissociation rates despite the higher breaking force observed for WT (Fig. 1d) suggest that the IDRs increase resistance to unzipping without altering the equilibrium dissociation pathway, thus acting as a source of molecular friction.

Analysis of the full probability distributions revealed further complexity. While unbound-state durations followed a single-exponential distribution (Supplementary Fig. 3a), the bound-state durations deviated from this, suggesting the presence of two kinetically distinct populations of Msn2–DNA bound complexes (Fig. 2d). This biphasic behavior was observed for both WT and DBD variants and appears to be specific to Msn2, as it was not observed for the previously studied Egr-1 DBD (Supplementary Fig. 3b). The rare, long-lived states may reflect slow conformational rearrangements within the Msn2–DNA complex, analogous to the bound-state structural heterogeneity reported for NF-κB on a single κB site[29], where distinct conformations interconvert over timescales comparable to or exceeding the overall residence time. Notably, while the rapidly dissociating subpopulation showed similar lifetimes for WT and DBD, the more stable subpopulation was longer-lived for the WT protein (Fig. 2e). These findings suggest that, in addition to promoting faster target association, the IDRs also stabilize a subset of tightly bound Msn2–DNA-bound complexes.

IDRs promote cooperative non-specific DNA binding by Msn2

During our analysis of the equilibrium binding experiments (Fig. 1), we discovered an unexpected phenomenon that provided further insight into IDRs' function. In some unzipping cycles, we observed binding events tens to hundreds of base pairs away from the canonical binding site (Fig. 3a and Supplementary Fig. 4a). These non-specific interactions, so called due to the absence of a canonical Msn2 motif, showed no correlation with motif similarity (Supplementary Fig. 4b), and were predominantly detected with WT Msn2, rarely with the DBD variant, and never in the absence of protein, indicating that the IDRs are critical

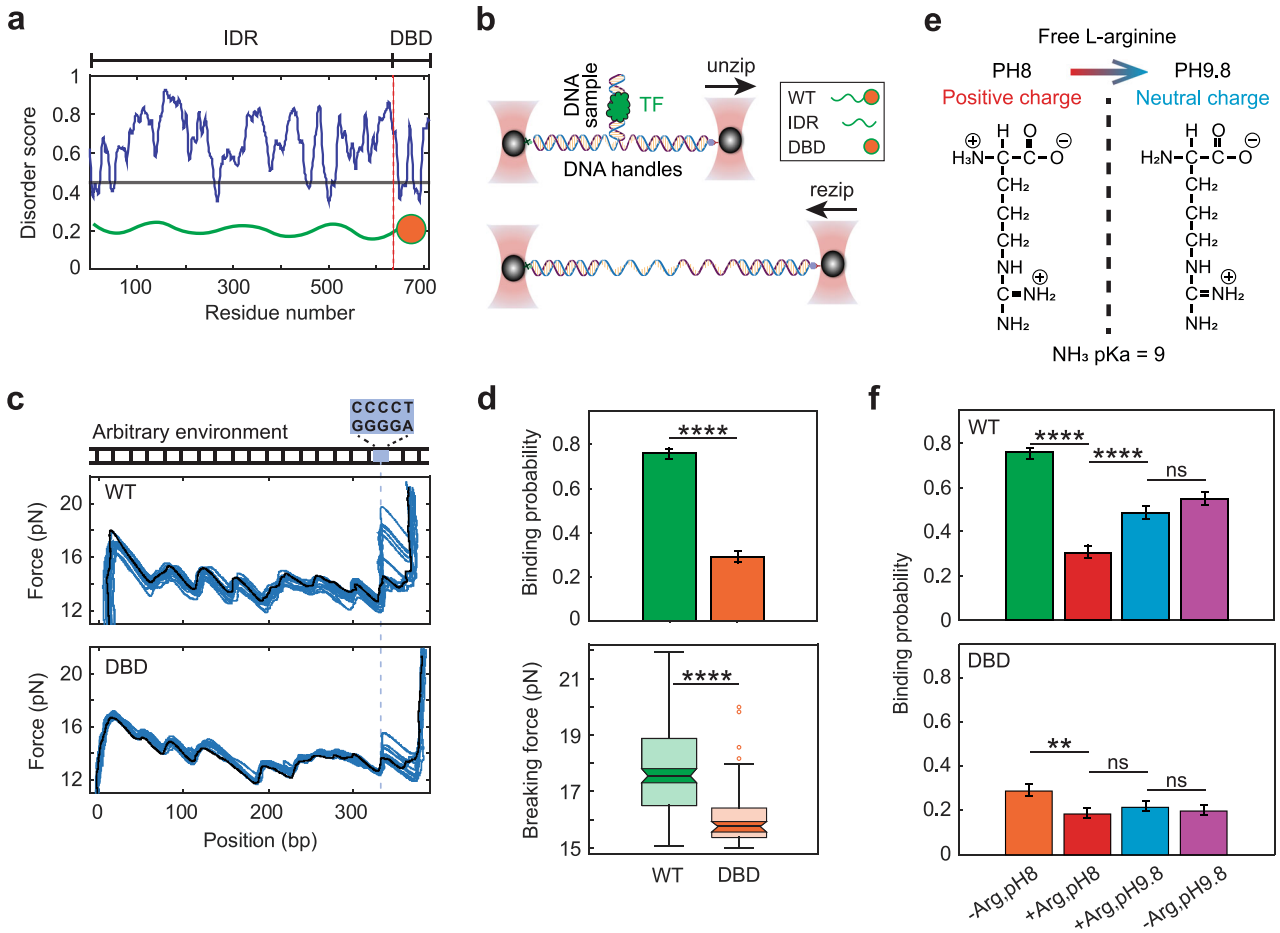

Fig. 1 | IDRs increase Msn2 binding affinity via charge-mediated interactions. **a** Disorder prediction profile of Msn2 calculated using IUPred3, highlighting the intrinsically disordered regions (IDRs) and DNA-binding domain (DBD). **b** Schematic representation of the DNA unzipping-rezipping experimental assay used to compare binding properties of wild-type Msn2 (WT), IDR-only (IDR), and DBD-only (DBD) variants. See also Supplementary Fig. 1. **c** Representative force-position traces from DNA unzipping experiments in the presence of 50 nM WT (top) and DBD (bottom) variants. The DNA construct contains an AT-rich sequence flanking the binding motif (position marked by dashed line). The black trace shows unzipping in a protein-free solution for reference. **d** Binding probability (total unzipping iterations $n_{WT} = 298$, $n_{DBD} = 289$) and breaking force (n_{WT}=226, $n_{DBD} = 83$) for WT and DBD variants. Binding probabilities are shown as the fraction of cycles with a binding event ± SE. Breaking forces are presented as box plots, showing the median (center line), interquartile range (IQR; box), whiskers extending to 1.5 × IQR, and notches denoting the 95% confidence interval of the median. ****$P < 0.0001$, two-sided χ^2 test and Student's t test, respectively. **e** Charge state of free L-arginine at pH 8.0 (left, positively charged) and pH 9.8 (right, neutral) illustrating the pH-dependent protonation of the amino group (pKa=9.0). **f** Binding probability under different electrostatic perturbation conditions for WT (top panel; total unzipping iterations $n_{-Arg,pH8} = 298$, $n_{+Arg,pH8} = 258$, $n_{+Arg,pH9.8} = 319$, $n_{-Arg,pH9.8} = 302$) and DBD variants (bottom panel; $n_{-Arg,pH8} = 289$, $n_{+Arg,pH8} = 309$, $n_{+Arg,pH9.8} = 339$, $n_{-Arg,pH9.8} = 342$). Conditions from left to right: "unperturbed" (similar to D), electrostatic screening with positively charged L-arginine at pH 8.0, neutralized screening with L-arginine at pH 9.8, and control at pH 9.8 without L-arginine. Data shown as the fraction of cycles with a binding event ± SE, **$P < 0.01$, ****$P < 0.0001$, two-sided χ^2 test. Exact P values for all statistical comparisons are provided in Supplementary Table 5. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-025-67217-2

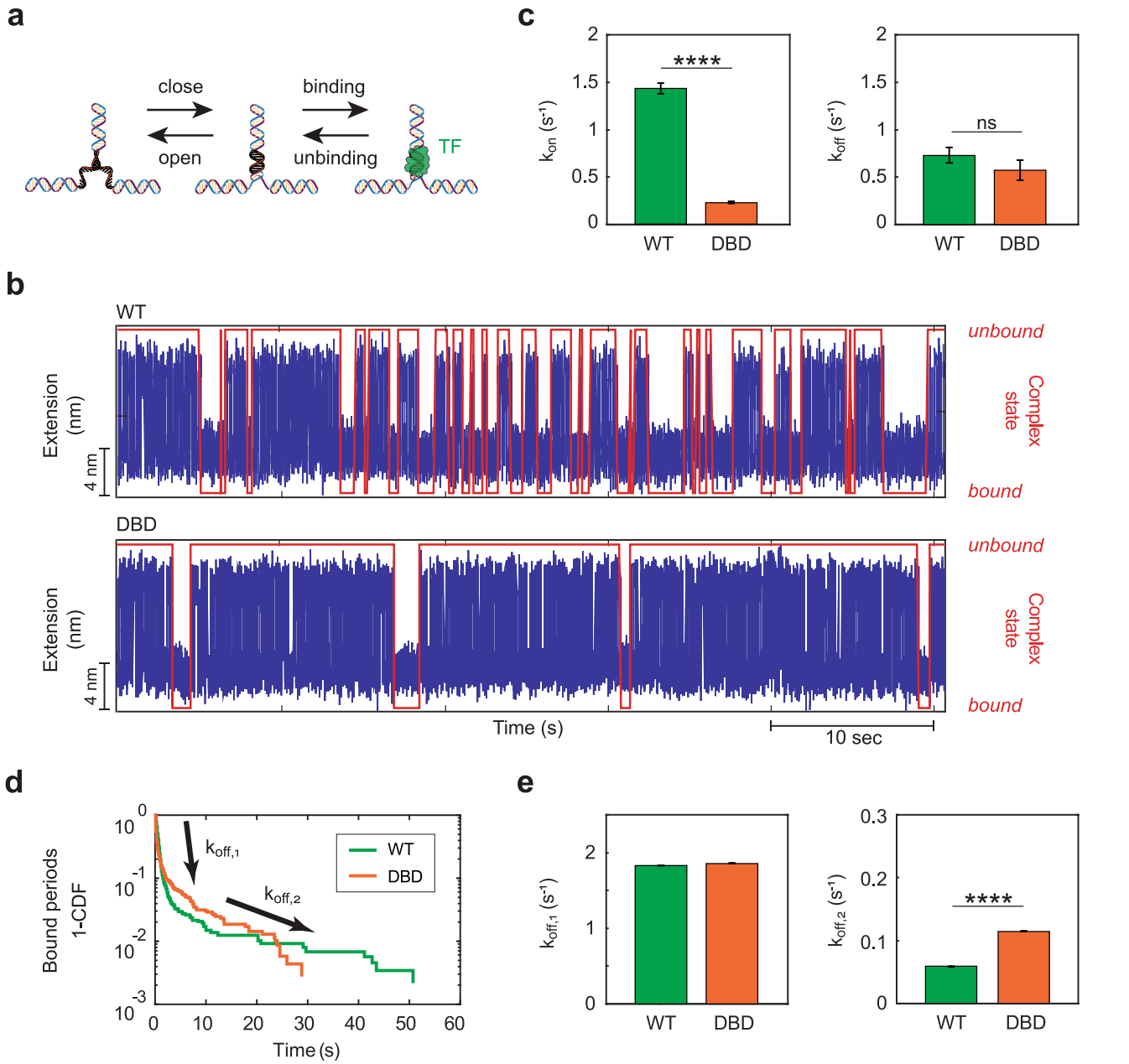

Fig. 2 | IDRs enhance the association rate of Msn2 to its binding motif. **a** Schematic illustration for the kinetic fluctuation suppression assay. After partial unzipping, the DNA construct undergoes rapid thermal fluctuations between "open" and "closed" states. When a transcription factor binds to its recognition site within the fluctuating segment, these fluctuations are suppressed. **b** Representative experimental traces for the kinetic fluctuation suppression assay at the presence of WT (top) and DBD (bottom). **c** Comparison of association rates (left panel; total period events $n_{WT} = 838$, $n_{DBD} = 642$) and dissociation rates (right panel; $n_{WT} = 892$, $n_{DBD} = 702$) for 10 nM Msn2 WT and DBD variants at the specific binding site. Data shown as mean ± SEM, $^{****}P < 0.0001$, two-sided permutation test with 10,000 resampling. **d** Cumulative distribution function (CDF) of bound state durations for both WT and DBD variants, revealing two distinct populations with different dissociation rates, $k_{off,1}$ and $k_{off,2}$. **e** Quantitative comparison of the two dissociation rates between WT and DBD variants, calculated from a double-exponential fit to the bound-state duration distributions (total period events $n_{WT} = 623$, $n_{DBD} = 307$). Data shown as mean ± SEM, $^{****}P < 0.0001$, two-sided permutation test with 10,000 resampling. Exact *P* values for all statistical comparisons are provided in Supplementary Table 6. Source data are provided as a Source Data file.

for their formation (Fig. 3b). However, unzipping experiments with the IDR variant showed no non-specific (nor specific) events.

To further characterize the nature of the complexes involved in the observed non-specific binding, we conducted electrophoretic mobility shift assays (EMSA) with a 51 bp DNA segment containing the canonical binding motif. For WT Msn2, we observed two distinct bands, indicating the formation of two nucleoprotein complexes with different mobilities (Fig. 3c, left). Quantification with a sequential binding model revealed that the slower-migrating, larger complex accumulates cooperatively at the expense of the smaller one (Fig. 3d, left; Supplementary Table 4), consistent with cooperative multimer formation. At very high Msn2 concentrations (≥250 nM), gel migration is significantly reduced, consistent with higher-order or aggregate-like assemblies. Thus, the complexes observed at the concentrations used in the single molecule experiments (50 nM) represent discrete assemblies of a small number of Msn2 molecules bound to individual DNA fragments rather than non-specific aggregation. In addition, while no binding was detected for the IDR-only variant in our optical

https://doi.org/10.1038/s41467-025-67217-2

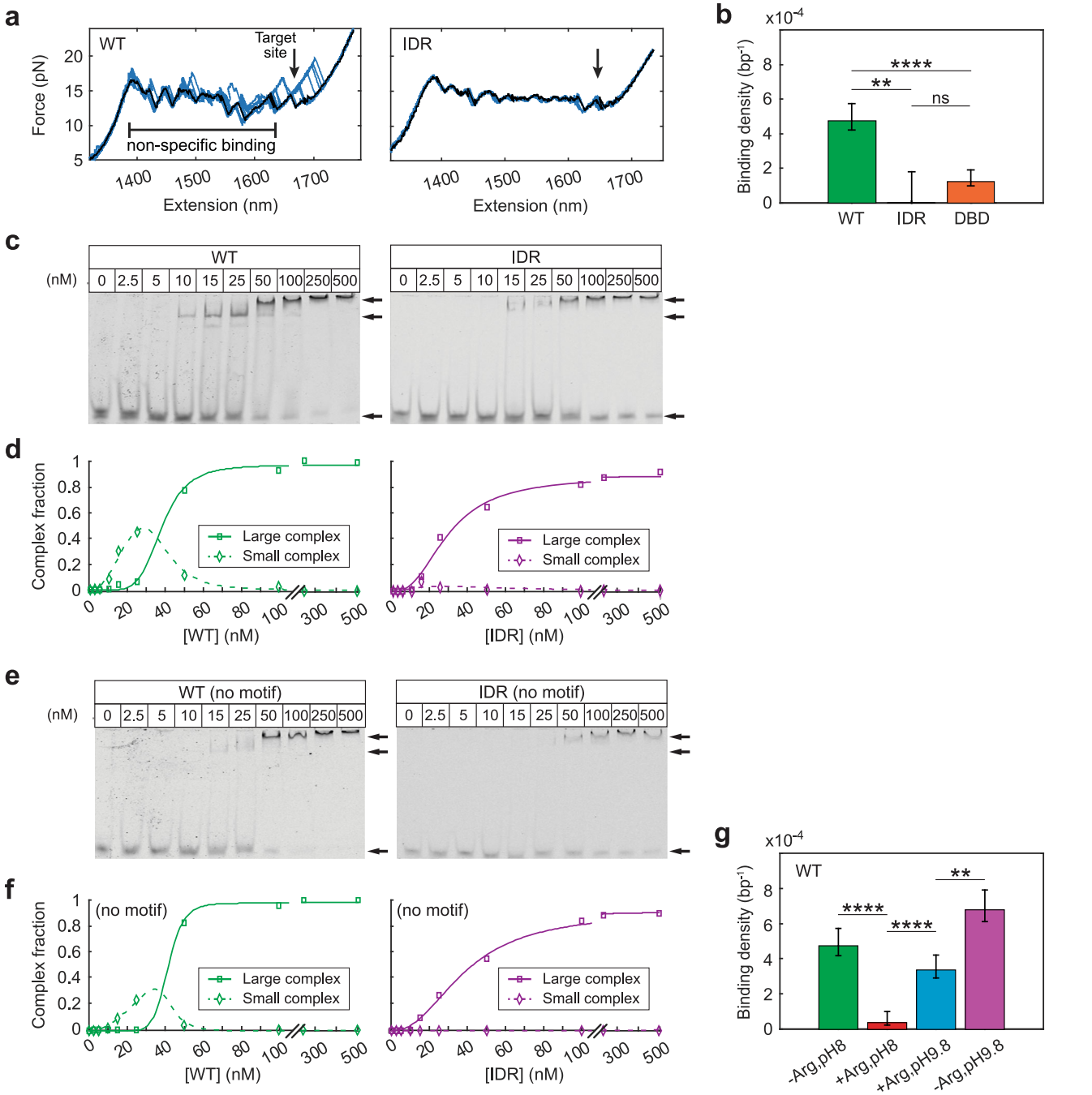

Fig. 3 | IDRs promote cooperative non-specific DNA binding by Msn2. **a** Representative force-extension traces from DNA unzipping experiments with Msn2 variants. The black trace was obtained in a protein-free solution for reference. Traces with WT Msn2 show non-specific binding events distant from the canonical binding site (marked by an arrow). **b** Non-specific binding density along the DNA construct prior to the binding site region, shown as the mean number of binding events, divided by segment length, ± SE (total position bins $n_{WT} = 9238$, $n_{IDR} = 868$, $n_{DBD} = 8959$). $^{**}P < 0.01$, $^{****}P < 0.0001$, two-sided χ^2 test. **c** EMSA using fluorescently labeled 51-bp DNA fragments containing the Msn2 binding motif, with increasing concentrations of Msn2 variants. In addition to the unbound DNA band at the bottom of the gel, two distinct bound complexes are observed in the upper bands (indicated by arrows). **d** Quantitative analysis of bound complex formation as a function of protein concentration for both complexes observed in the EMSA experiments. **e** EMSA with DNA sequences lacking the binding motif. **f** Quantitative analysis of complex formation for bound states observed in the motif-free EMSA experiments. **g** Non-specific binding density for the WT variant under various electrostatic perturbation conditions as described in Fig. 1 (total position bins $n_{-Arg,pH8} = 9238$, $n_{+Arg,pH8} = 8401$, $n_{+Arg,pH9.8} = 9424$, $n_{-Arg,pH9.8} = 9362$). Data shown as the mean number of binding events, divided by segment length, ± SE, $^{**}P < 0.01$, $^{****}P < 0.0001$, two-sided χ^2 test. Exact P values for all statistical comparisons are provided in Supplementary Table 7. Source data are provided as a Source Data file.

tweezers assay (Fig. 3a), EMSA revealed that it does engage DNA, forming only the higher-order complex and failing to produce the smaller one observed with WT (Fig. 3c, right; Fig. 3d, right). Since unzipping detects proteins that interfere with strand separation, this suggests that although the IDRs contribute to non-specific binding, they either do not bridge between the strands or do so weakly enough to prevent detection by unzipping fork blockage. Together, these findings support the idea that the IDRs mediate non-specific DNA interactions, but the DBD is required to stabilize these interactions and produce the unzipping signature.

We next asked whether these complexes require the presence of the canonical binding motif, and therefore replaced it with a random sequence (Fig. 3e). Removal of the motif completely abolished complex formation by the DBD variant (Supplementary Fig. 4c), whereas both WT and IDR were still able to form complexes. However, the substitution reduced the abundance of the small complex in WT while leaving the larger complex largely unchanged (Fig. 3e, f). These results confirm that the smaller complex depends on the DBD and reflects both specific and non-specific interactions. In contrast, the larger complex (IDR-dependent and DBD-stabilized) reflects cooperative non-specific interactions only. Interestingly, the DBD-only variant, which showed no binding in the absence of the motif, also produced two shifted bands in its presence (Supplementary Fig. 4c), suggesting that even motif-specific DBD–DNA interactions can generate multiple conformational or stoichiometric states. These may reflect distinct protein–DNA binding modes or DBD dimerization at the motif.

To test whether the formation of the IDR-dependent large complexes depends on charge-mediated interactions, we measured the density of non-specific Msn2 binding events observed in unzipping experiments under electrostatic perturbation with L-arginine. For WT Msn2, this treatment led to a clear reduction in non-specific binding (Fig. 3g). Increasing the pH to 9.8 partially reversed this effect, consistent with reduced electrostatic screening by the neutralized form of L-arginine. However, binding at pH 9.8 in the absence of L-arginine was substantially higher than at pH 9.8 with L-arginine, indicating that L-arginine retains an inhibitory effect under these conditions that cannot be explained solely by its net charge. This points to additional arginine-specific effects, such as direct interactions with DNA or protein. Notably, binding at pH 9.8 without L-arginine was somewhat enhanced relative to pH 8.0, suggesting that elevated pH alone does not impair non-specific binding and may even favor it. This contrasts with the behavior of site-specific interactions, where high pH reduced binding, and supports the idea that the two binding modes rely on distinct sets of interactions. While less frequent overall, non-specific binding by the DBD-only variant was also reduced by these perturbations (Supplementary Fig. 4d). Taken together, the EMSA and perturbation results show that IDRs support multiple, cooperative, non-specific contacts with DNA and that these interactions are at least partly mediated by electrostatics.

IDRs mediate Msn2 binding to single-stranded DNA

During inspection of the rezipping traces from our equilibrium unzipping experiments, we unexpectedly observed instances of hysteresis, i.e., instances where the force[29] during rezipping differed from the unzipping force at the same position (Fig. 4a and Supplementary Fig. 5a). These events, which occurred both near and far from the binding motif, reflect hindrance of DNA re-annealing and were rarely observed in the absence of protein, suggesting that Msn2 may bind to single-stranded DNA (ssDNA) exposed during unzipping. Hysteresis was significantly more pronounced with full-length Msn2 compared to the DBD and IDR variants (Fig. 4b), and complementary EMSA experiments showed formation of ssDNA-bound complexes for WT and IDR, but not when the IDRs were removed (Fig. 4c), suggesting that ssDNA binding is mediated primarily by the IDRs and might be stabilized by the DBD. As with double-stranded DNA (dsDNA) binding, ssDNA interactions were sensitive to L-arginine perturbation (Fig. 4d and Supplementary Fig. 5b). However, unlike dsDNA binding, ssDNA interactions were not rescued by raising the pH to 9.8 in the presence of L-arginine and remained low under all high-pH conditions, including in the absence of L-arginine. This suggests that IDR–ssDNA interactions rely on a distinct mechanism, sensitive to L-arginine but not governed solely by classical electrostatic screening.

A surprising result from the kinetic experiments provided additional support for these results. Under conditions where DNA undergoes thermal breathing (Supplementary Fig. 5c), the distribution of open-state lifetimes is typically well described by a single exponent (Supplementary Fig. 5d, left). However, in the presence of WT, but not the DBD variant, a second population of long-lived open states emerged (Supplementary Fig. 5d, center and right), indicating transient stabilization of the open state by protein binding to ssDNA. Together, these results reveal a previously uncharacterized ability of Msn2 to interact with ssDNA through its disordered regions.

IDRs enable a sequence-dependent target search on DNA

Our finding that IDRs promote non-specific DNA binding, together with the observation that they accelerate recognition of specific motifs by stabilizing an intermediate bound state in our kinetic analysis, suggests that IDRs enhance the ability of Msn2 to find its target sites, potentially through one-dimensional diffusion ("sliding") or other facilitated search modes. To test this model directly, we developed a new single-molecule assay, Search to Target Occupation (STO) (Fig. 5a). In this assay, the DNA construct is first fully unzipped, effectively eliminating the binding motif, and then incubated for 1 min in a solution containing Msn2. Following incubation, the construct is transferred using our laminar flow system into a TF-free channel, where no additional proteins can bind, and any dissociation is irreversible. Unzipping and rezipping cycles are then performed to monitor whether Msn2 can eventually reach and bind the target motif.

Remarkably, after incubation with WT Msn2, we observed specific binding events at the motif (Fig. 5b and Supplementary Fig. 6a, top) after a variable number of post-incubation unzipping cycles. No binding events were detected without prior incubation, indicating that Msn2 reaches the binding site by initially binding non-specifically elsewhere on the DNA and remaining associated through multiple unzipping and rezipping cycles (either by continuous binding or very short-range re-association, i.e., "hopping"). To rule out the possibility that STO binding events arose solely from extended binding to ssDNA during the incubation step, we repeated the assay with the DNA maintained in a closed dsDNA state during incubation (i.e. without prior unzipping, Supplementary Fig. 6c). This yielded similar STO probabilities (Supplementary Fig. 6d), showing that neither the presence of the motif nor the formation of ssDNA during incubation is necessary for successful site localization. Finally, when IDRs were compromised, either by truncation or screening, the number of specific binding events following incubation decreased sharply (Fig. 5c and Supplementary Fig. 6b). Together, these results provide direct evidence for an IDR-facilitated search mechanism that involves charge-mediated interactions.

Having established that IDRs facilitate target search, we next asked whether this mechanism could contribute to Msn2's selective promoter occupancy in vivo[3]. To test this, we replaced our standard ("arbitrary") DNA environment with a motif-free segment from the *Hap4* promoter, which is selectively bound by Msn2 in vivo[3]. The motif-containing segment was kept constant to avoid proximal flanking sequence effects near the target site[27]. Notably, WT Msn2 showed a markedly higher probability of STO binding on the *Hap4* sequence compared to the arbitrary one (Fig. 5d, e and Supplementary Fig. 7a). Moreover, specific binding events on the *Hap4* environment were typically detected in earlier post-incubation cycles (i.e., shorter times) (Supplementary Fig. 7b), indicating that the promoter environment increased both the probability and the apparent rate of target site localization.

The higher STO probability induced by *Hap4* could reflect enhanced initial non-specific binding, decreased dissociation during the scanning phase, or increased scanning efficiency. We therefore sought to determine which of these effects contributes to the measured outcome. Tracking non-specific binding events over successive unzipping–rezipping cycles revealed that their density remained approximately constant for both sequences (Supplementary Fig. 7d), suggesting that the dissociation rate during the scanning phase is

https://doi.org/10.1038/s41467-025-67217-2

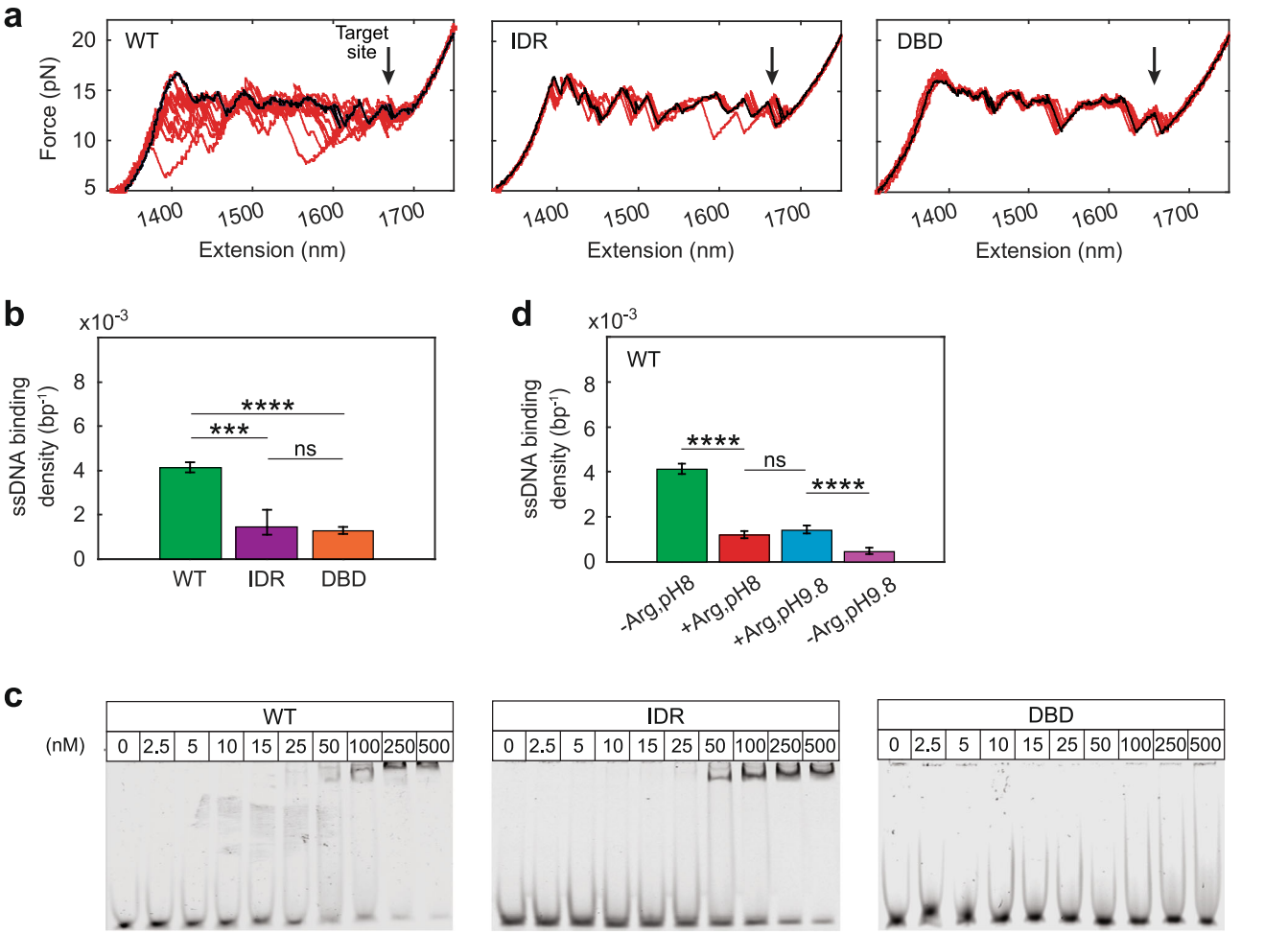

Fig. 4 | IDRs mediate Msn2 binding to single-stranded DNA. a Representative force-extension traces from DNA rezipping experiments with Msn2 variants. The black trace was obtained in a protein-free solution for reference. Traces with WT Msn2 show significant hysteresis due to non-specific binding to ssDNA at positions away from the binding site (marked by an arrow). **b** Quantification of ssDNA non-specific binding density, comparing Msn2 variants (total position bins $n_{WT} = 8959$, $n_{IDR} = 806$, $n_{DBD} = 8773$). Data shown as the mean number of binding events, divided by segment length, ± SE. ***$P < 0.001$, ****$P < 0.0001$, two-sided χ^2 test. **c** EMSA of fluorescently labeled ssDNA oligonucleotides with increasing concentrations of Msn2 variants. Bound complexes were observed for all variants except the DBD construct. **d** ssDNA non-specific binding density for the WT variant under different electrostatic perturbation conditions as described in Fig. 1 (total position bins $n_{-Arg,pH8} = 8959$, $n_{+Arg,pH8} = 8029$, $n_{+Arg,pH9.8} = 7068$, $n_{-Arg,pH9.8} = 8928$). Data shown as the mean number of binding events, divided by segment length, ± SE. ****$P < 0.0001$, two-sided χ^2 test. Exact P values for all statistical comparisons are provided in Supplementary Table 8. Source data are provided as a Source Data file.

insensitive to sequence, and thus cannot explain the higher STO probability for *Hap4*. In contrast, we observed that the *Hap4* promoter sequence induced significantly more non-specific binding than the arbitrary sequence, reflected in both the density of non-specific events (Fig. 5f) and the fraction of molecules showing such events (Supplementary Fig. 7c). Hence, the initial binding, referred to here as "non-specific" due to the absence of a canonical motif, is in fact sequence-sensitive. Supporting this interpretation, the density of non-specific binding events was non-uniform along the DNA sequence (Supplementary Fig. 7e). Notably, this density was uncorrelated with motif similarity (Spearman $r = -0.272$, $p = 0.132$ for *Hap4*), indicating that near-cognate motifs do not explain the sequence-sensitivity of non-specific binding. Finally, to test whether the scanning phase is also sequence dependent, we added L-arginine selectively to the TF-free channel, aiming to perturb IDR interactions during this phase without affecting initial non-specific binding. Under these conditions, both STO probability and detection times on the arbitrary sequence remained unchanged (Fig. 5e, Supplementary Fig. 7b). In contrast, the *Hap4* promoter exhibited a sharp reduction in STO probability and a shift toward longer detection times, reaching values comparable to those observed for the arbitrary sequence (Fig. 5e and Supplementary Fig. 7b). Similar effects were observed when IDR functionality was perturbed with increasing KCl (Supplementary Fig. 7b, f). These results suggest that while scanning can occur at a basal level with compromised IDRs activity, a sequence-specific enhancement is supported by IDRs–DNA interactions. Together, these findings support a role for IDRs in enabling promoter-specific search by promoting both initial DNA engagement and preferential exploration of favorable sequences.

Discussion

While canonical DBDs recognize short core sequence motifs, these motifs are typically degenerate and widely distributed throughout the genome. Yet in vivo, TFs occupy only a small fraction of these potential sites, suggesting that additional mechanisms confer functional specificity. This raises a fundamental question: how do TFs selectively bind their functional targets within a dense background of decoys? Motivated by in vivo studies showing that IDRs contribute to promoter specificity[3,5], we used single-molecule optical tweezers to investigate the molecular mechanisms underlying this selectivity, using the yeast TF Msn2 as a model. Our findings reveal a multifaceted role for IDRs in

https://doi.org/10.1038/s41467-025-67217-2

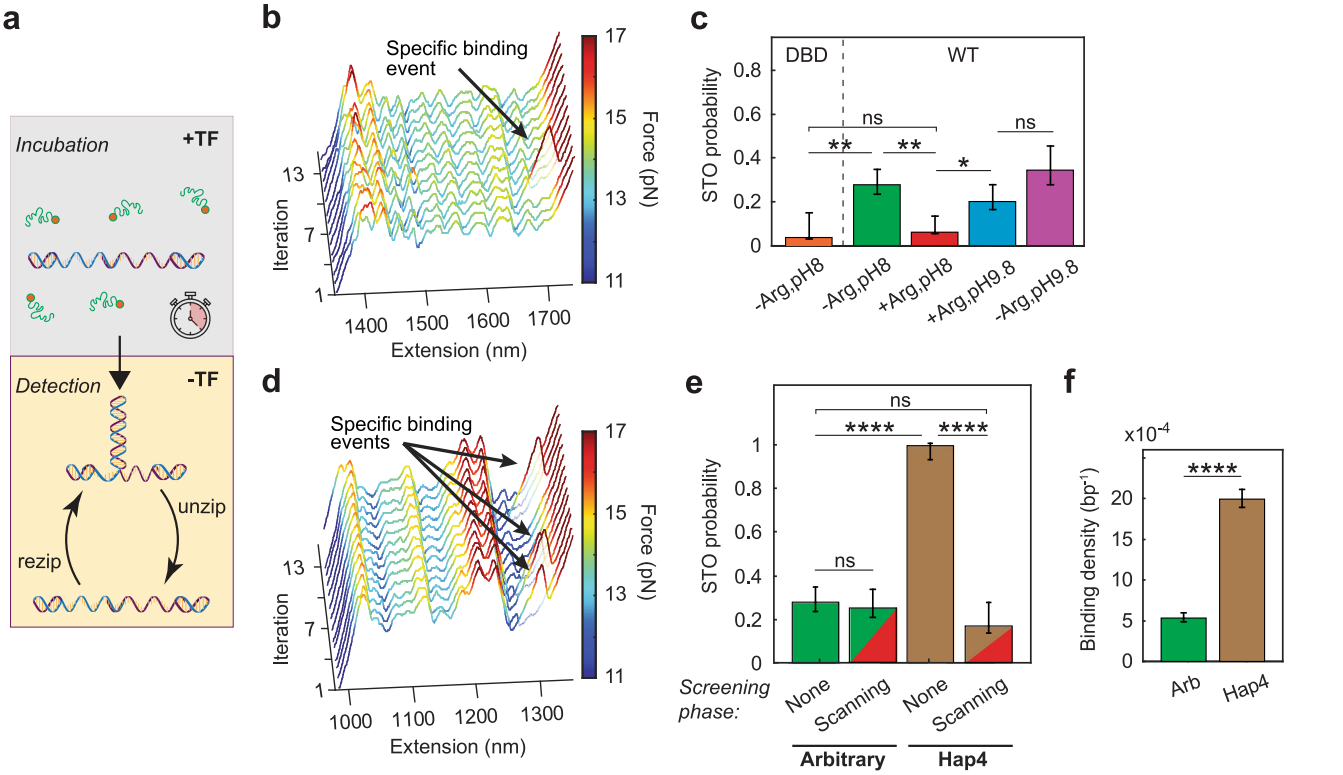

Fig. 5 | IDRs enable sequence-dependent target search on DNA. a Schematic of the Search-to-Target Occupation (STO) assay. During the incubation phase, fully unzipped DNA is exposed to transcription factors, allowing non-specific loading. The construct is then transferred to a protein-free channel for the detection phase, where repeated unzipping–rezipping cycles monitor whether proteins reach the binding site via 1D diffusion and/or other facilitated search modes. **b** Representative unzipping force-extension traces from the detection phase following a one-minute non-specific incubation with WT Msn2. A specific binding event at the target motif (arrow) appears during the fourth iteration. Each cycle lasts ~8 s. **c** STO probability, defined as the fraction of DNA constructs showing specific binding events following non-specific incubation. IDR functionality was assessed by comparing WT Msn2 to either the DBD variant or the WT protein under electrostatic perturbation conditions as described in Fig. 1 (total DNA molecules probed $n_{DBD} = 27$, $n_{-Arg,pH8} = 61$, $n_{+Arg,pH8} = 49$, $n_{+Arg,pH9.8} = 50$, $n_{-Arg,pH9.8} = 26$). Data shown as the fraction of cycles with a binding event ± SE, $^{*}P < 0.05$, $^{**}P < 0.01$, two-sided χ^2 test. **d** Representative force-extension traces during the detection phase using a DNA construct in which the flanking sequence was replaced with a segment from the *Hap4* promoter. Specific binding events occurred more frequently than in the arbitrary-sequence environment. **e** WT Msn2 STO probability comparison between different sequence environments flanking a constant binding site, with (split color bar) and without (solid color) electrostatic screening (50 mM L-arginine) applied during the scanning phase only. Left: arbitrary-sequence environment (green bar is same as -Arg,pH8 condition in C). Total number of DNA molecules probed $n_{None} = 61$, $n_{Scanning} = 51$. Right: *Hap4* promoter. Total number of DNA molecules probed $n_{None} = 50$, $n_{Scanning} = 30$. Data shown as the fraction of cycles with a binding event ± SE, $^{****}P < 0.0001$, two-sided χ^2 test. **f** Non-specific binding density for WT Msn2 with arbitrary sequence and *Hap4* promoter sequence (total position bins $n_{Arbitrary} = 40486$, $n_{HAP4} = 21390$). Data shown as the mean number of binding events, divided by segment length, ± SE. $^{****}P < 0.0001$, two-sided χ^2 test. Exact *P* values for all statistical comparisons are provided in Supplementary Table 9. Source data are provided as a Source Data file.

supporting efficient and selective DNA search, demonstrating that for Msn2, direct interactions between disordered regions and DNA play a central role in conferring sequence specificity through mechanisms beyond canonical motif recognition.

Our results support a model in which Msn2 engages with DNA through a stepwise process facilitated by its disordered regions (Fig. 6), which ultimately increases the association rate to the cognate motif by providing an additional binding pathway separate from direct binding from the solution. First, the IDRs mediate initial binding, allowing Msn2 to associate with DNA in the absence of a canonical motif. Once bound, Msn2 scans the DNA, ultimately leading to recognition and stable binding at the canonical AGGGG motif. IDRs enhance this search process, thereby promoting more efficient and selective occupation of target sites. A central insight from our study is that both initial binding and the subsequent DNA scanning are sequence sensitive. DNA derived from the *Hap4* promoter, which is selectively bound by Msn2 in vivo[3], enhances the formation of non-specific complexes and increases the probability and speed of site localization. These effects are abolished when the IDRs are removed. Site localization is also affected when the scanning phase is perturbed without affecting initial binding, demonstrating that features of the promoter environment modulate both the likelihood of initiating a productive search trajectory and the efficiency with which the target is reached. Thus, sequence-selectivity emerges not only at the recognition site itself, but throughout the entire search process, and is encoded in the dynamic interactions between Msn2's domains and the surrounding DNA. Notably, the observed sequence-selectivity, induced by replacing ~300 bp of flanking sequence while keeping the binding motif constant, reveals that search efficiency can be modulated over a short, ~100 nm range. This spatial scale is below the resolution of fluorescence-based single-particle tracking approaches, underscoring the unique ability of single-molecule mechanical assays to resolve localized sequence effects on transcription factor dynamics.

The proposed model aligns with the classic facilitated diffusion framework for transcription factor target search[12,30,31] and observations of 1D diffusion for various TFs[32–34], while introducing an additional layer by demonstrating the critical role of disordered regions in this process in the case of Msn2. An important question raised by our results is the precise mechanism by which Msn2 explores DNA once initially engaged. Because our assay involves a single DNA molecule, intersegmental transfer between different DNA molecules is excluded. Transient looping within the same DNA molecule could, in principle,

https://doi.org/10.1038/s41467-025-67217-2

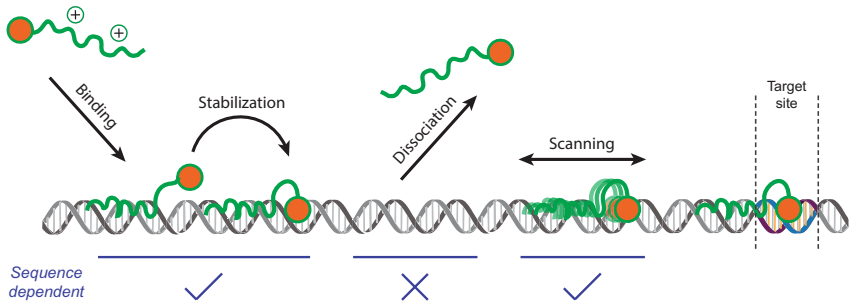

Fig. 6 | An emerging model of the IDRs-mediated searching mechanism of Msn2. Schematic representation of how Msn2 uses its intrinsically disordered regions (IDRs) to facilitate DNA binding and selective target localization. The process begins with (1) initial cooperative non-specific binding, mediated by electrostatic interactions between the positively charged IDRs and the DNA, followed by (2) stabilization via the DNA-binding domain (DBD). Binding, stabilization, or both contribute to sequence selectivity. Next, (3) one-dimensional scanning along the DNA, modulated by the IDRs in a sequence-sensitive manner. Finally, (4) the protein stably binds the specific recognition motif. This mechanism illustrates how IDRs tune both the kinetics and spatial specificity of transcription factor target search, thereby contributing to promoter selectivity.

permit transfer between distant segments. However, our DNA environment is only 300 bp in length (about twice the DNA persistence length) and is held under gentle flow, making extensive looping unlikely, though not impossible. We therefore favor a model in which scanning is dominated by one-dimensional diffusion, likely involving a combination of sliding and short-range hopping. In this view, the IDRs may stabilize sliding trajectories and/or promote rapid rebinding following transient dissociation. Notably, a recent theoretical study proposed that TFs with IDRs can enhance target search by first engaging DNA non-specifically through their IDRs and then dynamically sampling nearby sites via a flexible "octopusing" mechanism[35]. Similarly, in the context of protein–protein interactions, theoretical work suggested that disordered linkers may accelerate association by facilitating multivalent, polymer-like exploration of binding configurations, which may be maximized at an optimal IDR length[36]. Together, these theoretical frameworks and our single-molecule results support the idea that IDRs may act as kinetic enhancers of molecular search, raising the possibility that similar principles could apply across other protein–DNA or protein–protein systems. How broadly these principles apply to other TFs remains an open and important question.

We show that disordered and structured regions cooperate to enable stable DNA engagement by Msn2 away from the canonical binding motif. While the IDRs mediate non-specific interactions with DNA, these alone were insufficient to generate detectable unzipping signatures in our single-molecule assay. In contrast, full-length Msn2, which contains both IDRs and the DBD, was detected at non-motif sites. This suggests that the DBD, even outside its cognate sequence, provides additional stabilization that cooperates with the IDRs to form strand-engaging complexes. Rather than acting independently, the disordered and structured domains function as an integrated unit: the IDRs promote charge-mediated initial contact with the DNA, while the DBD reinforces these interactions, likely through weak hydrogen bonding or van der Waals contacts. Notably, while we refer to the initial binding as "non-specific," it exhibits clear sequence preference, as it was significantly increased for the *Hap4* promoter. This increase may arise from differential association by the IDRs, selective stabilization by the DBD, or both. Sequence-sensitive stabilization by the DBD could, in principle, arise from sites with high similarity to Msn2's motif. Motif similarity analysis (Supplementary Fig. 9) revealed only a small number of high-similarity sites (score >0.9) in both sequences. However, the distribution was modestly shifted toward higher similarity values in *Hap4*, consistent with a greater number of intermediate-similarity sites. In the case of the IDRs, several biophysical mechanisms could support such sequence selectivity. DNA sequences generate distinct electrostatic and mechanical landscapes[37,38] that can influence how IDRs conform around the DNA and facilitate scanning. Consistent with this, a bioinformatic comparison of the features of the two environments reveals that the GC-richer *Hap4* has a less negative electrostatic potential (EP), less negative propeller twist (ProT), somewhat smaller helical twist (HelT), and reduced fluctuations in ProT and minor-groove width (MGW), especially when the 100 bp proximal to the specific binding motif were compared (Supplementary Figs. 10 and 11). Several of these features correlate with GC content, and our analysis involves only a single sequence pair, so the results should be interpreted with caution. Nevertheless, a less negative EP is expected to lower electrostatic repulsion against the negatively charged IDRs and reduce the counter-ion release penalty during initial binding, while the other features suggest a more uniform, planar helical geometry that could reduce roughness for one-dimensional sliding by Msn2. Perturbation experiments helped clarify the molecular interactions through which IDRs contribute to the search process. The addition of 50 mM free L-arginine to the buffer, positively charged at physiological pH, led to a marked reduction in both non-specific binding and specific site engagement, in equilibrium and STO assays alike. For motif-specific binding, control experiments with KCl and pH manipulation support a strong electrostatic contribution. For non-specific binding, the effects of arginine and pH differed between dsDNA and ssDNA, and indicate additional, non-electrostatic contributions (see Supplementary Discussion). Notably, our findings highlight the critical role of charge-mediated interactions, likely arising from local positively charged residues within the overall negatively charged IDRs, in enabling an IDR-mediated search mechanism by facilitating initial binding and DNA association. However, they do not imply that electrostatics dominates the sequence-dependent aspects of the search, whose molecular basis may involve additional interactions, such as hydrophobic contacts, previously shown in vivo to contribute to promoter-specific occupancy[39]. In addition, while we did not test partial IDR truncations, in vivo studies have shown that promoter occupancy by Msn2 decreases gradually with IDR shortening, suggesting a distributed contribution of weak binding elements[3].

Our observation that Msn2 can bind to ssDNA through its IDRs, supported by the hysteresis observed during rezipping experiments and the extended open periods in thermal fluctuation measurements, represents a novel finding with potential functional implications. Transcriptionally active regions often contain transient ssDNA[40], most notably within transcription bubbles generated during RNA polymerase passage and in R-loop structures, where RNA: DNA hybrids leave the non-template strand single-stranded. These structures are prevalent across eukaryotic genomes and can influence transcriptional regulation[41]. The ability of IDRs to engage with them could potentially

contribute to transcription factor recruitment to active genes or influence the dynamics of transcriptional regulation. Recent work[42] demonstrated that the intrinsically disordered C-terminal domain of linker histone H1 exhibits preferential binding to single-stranded nucleic acids and can undergo liquid-liquid phase separation upon binding. This suggests that IDR-mediated interactions with single-stranded nucleic acids may represent a common functional modality across diverse DNA-binding proteins. Notably, the flexible and dynamic nature of ssDNA may promote non-electrostatic contacts, such as hydrogen bonding and potentially π–π or cation–π interactions involving aromatic or basic residues. This is supported by structural analyses of RNA–protein complexes[43], suggesting that similar interactions could contribute to IDR engagement with single-stranded nucleic acids.

While our current study focuses on DNA binding in a purified system, TFs must navigate a considerably more complex chromatin environment in vivo. Nucleosomes represent a major barrier to TF binding, potentially occluding binding sites and modulating the search process, as recently shown for the *Drosophila* GAGA factor[34]. Our previous work showed that nucleosomes modulate the energy landscape for TF binding through multiple mechanisms, including site accessibility and nucleosome dynamics[44–46]. Others have shown that the presence of nucleosomes can alter TF-DNA interactions, modulating TF-DNA recognition beyond simple steric occlusion[47]. Our results suggest two ways in which IDR–DNA binding could operate in this chromatin context. First, during scanning along DNA, the ability of Msn2's IDRs to engage histone components, particularly the flexible and acidic histone tails, through electrostatic and hydrophobic contacts, may allow the factor to move across or around nucleosomes rather than being blocked by them. Second, if the target motif is occluded within a nucleosome, non-specific binding by the IDRs to DNA and histone surfaces could anchor Msn2 nearby, enabling it to wait for transient unwrapping events that expose the site. Both effects could improve search efficiency in chromatin compared to a mechanism relying solely on specific DBD–DNA recognition. In line with these possibilities, in vivo studies have shown that IDRs direct Msn2 to promoters with "fuzzy" nucleosome architectures[5].

Finally, given the prevalence of intrinsic disorder in eukaryotic TFs[48], the search mechanism identified here may represent a fundamental principle that has been repeatedly exploited throughout evolution. While DBDs typically change through discrete mutational events that alter specific base contacts, IDRs can evolve more rapidly due to their tolerance for a wider range of sequence variations, potentially accelerating the evolution of gene regulatory networks[49]. Our findings suggest that modifications to IDRs could tune not only the activation potential of TFs but also their DNA search properties and target specificity. If so, the disordered architecture of TFs may be as critical to their regulatory specificity as their structured DBD, challenging conventional views of sequence-specific DNA recognition.

Methods

Expression and purification of Msn2 proteins

All three Msn2 proteins were expressed using the pET-24(b+) vector in *E. coli* BL21 (DE3) cells. The details of the constructs are presented in Supplementary Table 1. The WT and IDR (N-terminal His6-tagged) were purified from inclusion bodies (IBs), whereas the DBD (N-terminal His6-MBP-tagged) was purified from the soluble fraction.

For the WT and IDR, the bacteria were grown in 4l LB medium containing 50 μg/ml kanamycin, 0.5% glucose, and 4 g/l serine at 37 °C (250 rpm). The cultures were induced with 1.0 mM IPTG at OD_{600} 0.8 and incubated overnight under the same conditions. The cells were harvested by centrifugation (3000 × *g*, 30 min, 4 °C), and the cell pellets were suspended in ice-cold lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, Roche cOmplete protease inhibitor cocktail tablet (1 for each 15 ml). After sonication (70% amplitude, pulse on 2 s, pulse off 10 s, 60 cycles, on ice) and centrifugation (48,200 × *g*, 30 min at 4 °C), the precipitate was mixed with 60 mM EDTA, 6% Triton and 1.5 M NaCl, homogenized, and incubated for 30 min at 4 °C to solubilize membrane debris. The resulting suspension was centrifuged (48,200 × *g*, 30 min at 4 °C) and the IB precipitate was re-suspended in 100 mM Tris-HCl pH 8.0 and 1 mM EDTA, homogenized, and centrifuged using the same conditions. This IB wash was repeated 2 times. Finally, the IB pellet was re-solubilized in 50 mM Tris-HCl pH 7.5, 10 mM imidazole, 6 M GdmCl, 100 mM DTT and 10 mM EDTA for more than 2 h at 4 °C. The supernatant was collected by centrifugation (48,200 × *g*, 30 min at 4 °C) and the buffer was exchanged to 50 mM Tris-HCl pH 7.5, 6 M GdmCl, and 10 mM imidazole, using a HiPrep desalting column (26/10, GE) to remove DTT and EDTA. The protein-containing fraction was then loaded on a HisTrap Ni-NTA column (5 ml, Cytiva) and the protein was eluted with 50 mM Tris-HCl, pH 7.5, 6 M GdmCl, and 0.5 M imidazole. To remove the His6-tag, the eluted protein was supplemented with HRV3C protease (0.15 mg/ml) and the solution was first dialyzed against 125 volumes of cleavage buffer (50 mM Tris-HCl, 300 mM KCl, 300 mM L-Arg and 1 mM DTT, measured pH 7.5) for 1 h (buffer was changed after 30 min) at 25 °C and then against 200 volumes for overnight at 4 °C. GdmCl and imidazole powders (to final concentrations of 6 M and 10 mM, respectively) were added to the cleavage reaction and loaded on the HisTrap column. Cleaved protein was obtained in the flow-through, while the His6-tag and HRV3C protease were retained on the column. The cleaved protein was then reduced with 30 mM DTT, and further purified either by size-exclusion chromatography (SEC, Superdex 200 Increase 10/300 GL, Cytiva) and the protein-containing fractions were exchanged to storage buffer (20 mM Tris-HCl pH 8 and 150 mM KCl) using a HiTrap desalting column (5 ml, Cytiva), or passed it through a reverse-phase HPLC Jupiter C4 column and then dissolved in the storage buffer. Purity of the proteins was confirmed by SDS-PAGE (Supplementary Fig. 1). The protein concentration was determined using a predicted extinction coefficient of ε_{280} = 24,870 M^{-1} cm^{-1} (both for WT and IDR) and stored at −80 °C.

For the DBD, bacteria were grown in 4l LB medium containing 50 μg/ml kanamycin and 0.5% glucose at 37 °C (250 rpm). The cultures were induced at OD_{600} 0.8 with 0.5 mM IPTG and incubated at 25 °C (250 rpm) for an additional 8 h. The cells were harvested by centrifugation (3000 × *g*, 30 min, 4 °C), and cell pellets were suspended in ice-cold lysis buffer (50 mM Tris-HCl, pH 8, 10 mM EDTA, Roche complete protease inhibitor cocktail tablet (1 for each 15 ml). After sonication (70% amplitude, pulse on 2 s, pulse off 10 s, 60 cycles, on ice) and centrifugation (48,200 × *g*, 30 min at 4 °C), the supernatant was collected and reduced with 50 mM DTT on ice for 1 h. In the next step, the supernatant was loaded on an amylose column (resin from NEB, packed in GE XK 16/40 column), washed with 20 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM CaCl2 and the desired protein was eluted with 20 mM Tris-HCl pH 7.5, 50 mM NaCl, 1 mM CaCl2 and 10 mM maltose. To remove the DNA impurities, the protein was first reduced with 50 mM DTT on ice for 1 h, then diluted 5-fold with 15 mM Tris-HCl pH 7.3, 2 mM EDTA, and loaded on a DEAE FF ion-exchange column (5 ml, Cytiva). The flow-through (containing most of the protein) was collected, concentrated and then the buffer was exchanged to 20 mM Tris-HCl pH 7.5, 10 mM imidazole and 1.5 M NaCl, using the HiPrep desalting column. Afterwards, the solution was loaded on a 5 ml HisTrap column, washed with 10 column volume (CV) of 20 mM Tris-HCl pH 7.5, 25 mM imidazole and 0.5 M GdmCl, re-equilibrate with 20 mM Tris-HCl pH 7.5, 10 mM imidazole and 1.5 M NaCl, and finally eluted with 20 mM Tris-HCl pH 7.5, 0.5 M imidazole, and 1.5 M NaCl using a linear a gradient from 0 to 100% over 5 min (3 ml/min). The protein fractions were pooled, and the buffer was exchanged to 20 mM Tris-HCl, pH 7.5 and 100 mM NaCl using the HiTrap desalting column. To remove the MBP-His6-tag, the protein solution was incubated with 0.15 mg/ml HRV3C protease for 2 h at 25 °C and passed it through the reverse-phase HPLC Jupiter C4 column. The pure lyophilized DBD (as

confirmed by SDS-PAGE, Supplementary Fig. 1) was dissolved in storage buffer (20 mM Tris-HCl, pH 7.5 and 100 mM NaCl), the concentration was measured using bicinchoninic acid (BCA) assay (because DBD has a computed extinction coefficient of ε_{280} = 0 M^{-1} cm^{-1}) and stored at −80 °C.

Single-molecule experiments

The constructs for single-molecule experiments were generated as previously described[28,46] with some changes. All constructs were assembled by ligation of several DNA segments (Supplementary Fig. 8 and Supplementary Table 19): A pair of DNA "handles", an "upstream environment" segment (UE, ~300 bp), a "binding region" segment (BR), and a short "hairpin" segment (HP). For the kinetics experiments in Fig. 2, since the DNA is unzipped up to the binding site itself, an additional "downstream environment" segment (DE, ~250 bp) was inserted between the BR and HP, to create a downstream flanking sequence similar to the upstream one. The two ~2000 bp DNA "handles" were prepared as previously reported[28,45], each incorporating a specific tag (biotin or digoxygenin) on one end and a complementary ssDNA overhang on the other, and were then annealed. The UE was amplified from the *Cga* gene promoter using mouse genomic DNA as a template or, for the experiments in Fig. 5, from the *Hap4* gene promoter using *S. Cerevisiae* genomic DNA as a template. Primers are listed in Supplementary Table 15. PCR products were digested with DraIII-HF (R3510L; NEB) according to the manufacturer's instructions and purified using a QIAquick PCR Purification Kit (28106; QIAGEN). BR for the experiments described in Supplementary Fig. 2a is a ~600 bp segment from the *Hor7* gene promoter that includes four Msn2 binding motifs and was produced via PCR with the pET24b-Hor7 plasmid (kindly provided by Dr. H. Hofmann) as a template (primers listed in Supplementary Table 15). The product was digested with SFI-I (R0123S; NEB). In all other experiments, BR is a 60 bp AT-rich sequence flanking a single Msn2 binding motif and was produced by annealing two ssDNA oligonucleotides (Supplementary Table 16), designed with overhangs at their ends in preparation for their ligation assembly. Complementary ssDNA oligos were first phosphorylated using T4 Polynucleotide Kinase (M0201L; NEB), according to the manufacturer's instructions, mixed in a 1:1 molar ratio in T4 DNA Ligase Reaction Buffer (B0202S; NEB), incubated at 90 °C for 5 min, and cooled down slowly to room temperature. BR for the Egr1-based experiment, shown in Supplementary Fig. 3b, is a 68 bp sequence designed with a single Egr1 binding motif along with a 5 bp native flanking segment from the *Lhb* gene promoter, and was produced in a similar annealing method. DE was amplified from the *Cga* gene promoter by PCR with mouse genomic DNA (primers listed in Supplementary Table 15) and digested with DraIII-HF (R3510L; NEB). HP is a 27 nt ssDNA molecule that folds into a hairpin with a 10 bp stem (Supplementary Table 16), and contains a 3 nt overhang to facilitate ligation (TGC for the kinetic experiments, CTA for all other experiments).

To assemble the full construct, the annealed handles were ligated to UE in a 1:5 reaction overnight with T4 DNA ligase (M0202L; NEB), and the ligation product was gel-purified. Separately, BR was ligated to HP in a 1:5 reaction overnight with T4 DNA ligase (M0202L; NEB). Handles-UE and BR-HP were then ligated in a 1:5 reaction for 30 min at room temperature, using Rapid ligase (C671B; Promega). For the kinetic experiments (Fig. 2), DE was ligated to HP in a 1:5 reaction overnight with T4 DNA ligase (M0202L; NEB), and then a three-piece ligation with Handles-UE, BR and DE-HP, in a 1:5:25 ratio, was performed for 30 min at room temperature, using Rapid ligase (C671B; Promega).

The full construct was incubated for 15 min on ice with 0.8 µm polystyrene beads (Spherotech), coated with anti-Digoxigenin (anti-Dig) molecules. The reaction was then diluted 1000-fold in binding buffer (BF): 20 mM Tris·HCl pH 8 (or 9.8, as specified), 5 mM $MgCl_2$, 1 mM TCEP, 3% v/v glycerol and 0.01% BSA, 0.001% Tween20, 1uM EDTA, 2uM ZnCl, 150 mM KCl (unless otherwise indicated), and (when indicated) 50mM L-arginine. The binding buffer for Egr1 experiments was composed of 10 mM Tris·HCl, pH 7.4, 150 mM NaCl, 1.5 mM $MgCl_2$, 1 mM DTT, 3% v/v glycerol, and 0.01% BSA. Tether formation was performed in situ (inside the optical tweezers' experimental chamber) by trapping a DNA-bound anti-Dig bead in one trap, 0.9 µm streptavidin-coated polystyrene beads in the other trap, and bringing the two beads into proximity to allow binding of the biotin tag to the streptavidin-coated bead.

All the experiments with the Msn2 variants were conducted at 50 nM protein concentration, except for the kinetics experiments, which were conducted at 10 nM. Experiments with Egr1 were conducted at 4 nM.

Optical tweezers

Experiments were performed in a custom-made dual-trap optical tweezers apparatus, as previously reported[27,28,50]. Briefly, the beam from an 852 nm laser (TA PRO, Toptica) is coupled into a polarization-maintaining single-mode optical fiber. The collimated beam out of the fiber is split by a polarizing beam splitter (PBS) into two orthogonal polarizations, each directed into a mirror and combined again with a second PBS. One of the mirrors is mounted on a nanometer-scale mirror mount (Nano-MTA, Mad City Labs). An X2 telescope expands the beam while imaging the plane of the mirrors into the back focal plane of the focusing microscope objective (Nikon, Plan Apo VC 60X, NA/1.2). Two optical traps are formed at the objective's focal plane, each by a different polarization, and with a typical stiffness of 0.3–0.5 pN/nm. The light is collected by a second, identical objective, and the two polarizations are imaged onto two Position Sensitive Detectors (First Sensor), after being separated by a PBS. The instantaneous position of the beads relative to the center of their traps is determined by back focal plane interferometry[51]. Calibration of the setup is done by analysis of the thermal fluctuations of the trapped beads[52], which are sampled at 100 kHz. Experiments were conducted using a 5-channel laminar flow cell (Lumicks), which was passivated following a published protocol[53] with some modifications. Briefly, we washed the chamber twice by flushing alternately 1 M NaOH and 1% Liquinox for 10 min each. Casein (1%) was sonicated, filtered, diluted to 0.2%, and flushed into the chamber, where it was incubated for 40 min. After the incubation, the system was washed with binding buffer.

Data analysis

The fundamental analysis of the acquired data followed the same methodology previously described[28,46]. Briefly, data signals were acquired at 2500 Hz and stored for analysis using Matlab (MathWorks) scripts. The data was converted into force and extension vectors using the experimentally determined calibration parameters. When indicated (Fig. 1 and Supplementary Fig. 2) the position (or number of bp unzipped) was calculated by subtracting the stretching of the dsDNA and ssDNA parts of the construct, which we calculate using extensible-worm-like-chain (eWLC)[54] and worm-like-chain (WLC) models, respectively.

Motif-specific binding experiments. Motif-specific binding events were identified as events where the maximal force *Fi* at the nominal binding site position (±5 bp) exceeded a threshold force. This threshold was determined as the average maximal force at the binding site position from multiple traces taken in a protein-free solution, plus three times the standard deviation of these force values. When applying the threshold criteria to the data obtained from the protein-free solution, no binding events were detected. The binding probability was calculated as n/N, where n is the number of times a binding event was identified, and N is the total number of cycles. The uncertainty in the binding probability was calculated with the adjusted Wald interval of binomial proportions[55], and the significance of differences

was assessed using a two-sided χ^2 test. The breaking force was calculated as the mean Fi for the cycles where binding was detected. The significance of the difference between bound complex breaking forces was evaluated with a two-tailed Student's t test. Differences were considered statistically significant when $p < 0.05$.

Non-specific binding experiments. This experiment aims to detect non-specific binding events to either dsDNA or ssDNA, occurring over 20 bp away from the protein binding site motif. The maximal forces during unzipping iterations at a protein-free solution were measured throughout the DNA motif-flanking environment with 10 bp bins. This allowed us to set a threshold on the maximal force expected in the absence of a protein in a position-dependent manner, as three times the standard deviation above the averaged maximal force in each bin. The force data of each unzipping iteration, from experiments in the presence of the protein, was then compared, per bin, to the calculated threshold of that bin and marked as a dsDNA non-specific binding event if exceeded. Position-resolved non-specific binding density (Supplementary Fig. 7e) was calculated as the average number of bound events in each bin, divided by the bin size. The overall binding density was calculated by averaging over all the bins. Similarly, a minimal force threshold was calculated from the minimal force data of the rezipping iterations, to form a threshold on the minimal force expected in the absence of a protein and identify ssDNA non-specific binding events during rezipping iterations. The ssDNA binding density was calculated in a similar manner to the dsDNA binding density. To avoid misidentified events at positions that are particularly noisy in terms of their maximal force, unzipping/rezipping traces taken in protein-free solution were also verified against the calculated thresholds per position bin. In case one of the tested DNA molecules showed above/or below-threshold behavior in a protein-free solution, the relevant position bins were excluded from the binding density calculation of the corresponding molecule measurements that were held in the presence of the protein. Finally, the "background" density calculated in a protein-free solution was subtracted. The significance of differences between non-specific binding densities was evaluated with a two-sided χ^2 test. Differences were considered statistically significant when $p < 0.05$. In addition to this position-resolved method, an alternative method of quantifying non-specific binding was used in Supplementary Fig. 7c. The non-specific binding was calculated as the percentage of DNA molecules showing any non-specific binding event out of the total number of tested molecules. The significance of differences between these binding probabilities was evaluated with a two-sided χ^2 test. Differences were considered statistically significant when $p < 0.05$.

Search to target occupation (STO) experiments. The experiment aims to identify motif-specific binding events occurring in a protein-free solution following non-specific binding of the protein to the DNA construct. For that, separate channels (one protein-free channel and the second containing the variant under test) in our microfluidic laminar flow chamber (Lumicks) were utilized in a three-phase experiment. First, the DNA construct was fully unzipped in the protein-free channel. Next, it was translocated to the channel where the protein was present. At that stage, the unzipped construct had no effective binding site, hence proteins could only bind to it non-specifically. A 60-second incubation period was provided to allow stabilization of the non-specific binding process. Lastly, the construct was moved back to the protein-free chamber channel, where 20 unzipping-rezipping iterations were performed to probe for binding events. STO probability was calculated as the fraction of DNA constructs showing any specific binding event following non-specific incubation. The significance of differences between STO probabilities was evaluated with a two-sided χ^2 test. Differences were considered statistically significant when $p < 0.05$.

Kinetics experiments. Dissociation and pseudo–first-order association rates of the protein-DNA complex were calculated as previously described[28], with some changes as detailed below. Briefly, fixing the positioning of the unzipping fork in the vicinity of the protein's binding motif allows identifying protein binding by probing thermal fluctuations between "closed" (dsDNA) and "open" (two ssDNA) states, which are suppressed upon TF binding. After identifying closed and open stats using the HAMMY algorithm[56,57], their durations were determined. Closed states longer than a predetermined threshold, chosen to minimize the overall misidentified states, were classified as "bound" states. Non-bound periods above 15 ms were defined as "unbound" states. Association and dissociation rates were calculated as the reciprocal of the average unbound and bound states duration, respectively. Statistical significance of differences in kinetic rates was assessed using a two-sided non-parametric permutation test (10,000 iterations). In each iteration, dwell periods were randomly reassigned between the two condition groups to simulate the null hypothesis, and the rate difference was calculated. P-value was defined as the proportion of permuted rate differences with a magnitude greater than or equal to that of the observed difference. Dissociation rates for the two populations observed in Fig. 2c were calculated by fitting the bound periods' distribution with a double exponential function of the form $p \cdot \exp(-k_{off1}t) + (1-p) \cdot \exp(-k_{off2}t)$. Significant difference between the kinetic rates was assessed using a two-sided non-parametric permutation test (10,000 iterations). Differences were considered statistically significant when $p < 0.05$. As each test involved a single predefined comparison between two conditions, no correction for multiple comparisons was applied.

Electrophoretic mobility shift assays (EMSA)

A 51 bp dsDNA probe labeled with 5′ IRD800 was prepared by annealing 20 μM of each complementary oligonucleotide (IDT; sequences in Supplementary Table 17) in STE buffer (100 mM NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM EDTA) in a final volume of 100 μL. The mixture was heated to 95 °C for 5 min and then slowly cooled to room temperature overnight. Residual single-stranded DNA was removed by exonuclease treatment, and the resulting dsDNA was purified by ethanol precipitation. Binding reactions were performed in a 20 μl solution containing EMSA reaction buffer (20 mM Tris-HCl, pH 8, 150 mM KCl, 5 mM $MgCl_2$, 1 mM TCEP, 1 μM EDTA, 2 μM $ZnCl_2$), 1 nM DNA and 2.5-500 nM of the purified Msn2 variants (WT, IDR, or DBD). Reactions were incubated at room temperature for 30 min and then resolved on a non-denaturing 6% polyacrylamide gel prepared and run in 0.5× TBE buffer (45 mM Tris-borate, 1 mM EDTA). Samples were loaded directly onto the gel, which was run at 100 V for 60 min at room temperature.

Gels were scanned using an Odyssey DLx imaging system (LI-COR), and band intensities were quantified using Image Studio Lite software. Signal intensities corresponding to the bound and unbound DNA fractions were extracted. Quantification was performed using a thermodynamic equilibrium model with two sequential binding steps. Under the assumption that the total DNA concentration is much lower than the total transcription factor (TF) concentration, the fractional occupancies were calculated as $c_1 = \frac{1}{D}\frac{c^{n_1}}{k_1^{n_1}}$ and $c_2 = \frac{1}{D}\frac{c^{(n_1+n_2)}}{k_1^{n_1}k_2^{n_2}}$, where $D = 1 + \frac{c^{n_1}}{k_1^{n_1}} + \frac{c^{(n_1+n_2)}}{k_1^{n_1}k_2^{n_2}}$, c is the free TF concentration, k_1 and k_2 are the dissociation constants for the first and second binding steps, and n_1 and n_2 are their respective Hill coefficients. Model parameters were determined by non-linear least-squares fitting to the experimental data using MATLAB's Curve Fitting Toolbox, with initial parameter estimates based on observed half-maximal binding concentrations. Goodness of fit was evaluated using the coefficient of determination (R^2) and residual analysis.

DNA features analysis

Motif similarity was quantified relative to the position weight matrix (PWM) of Msn2 (JASPAR[58], MA0341.1). For each 10-bp window, all overlapping 5-bp segments were scored against the PWM on both the forward and reverse-complement strands. For each segment, the log-likelihood under the PWM was computed and normalized between the theoretical minimum and maximum possible scores, yielding values between 0 and 1. The mean of all normalized scores within the window was then taken as the motif similarity of that window. DNA structural and dynamic features were predicted from sequence using deepDNAshape[59,60], from which we extracted MGW, ProT, roll, HelT, and tilt, together with their estimated fluctuations (standard deviations). We also used deepDNAshape to calculate the EP. For each feature, Supplementary Fig. 10 shows raw position-resolved values, as well as cumulative distribution functions, for the "arbitrary" sequence and *Hap4*. Statistical significance of distributional differences was tested using the two-sample Kolmogorov–Smirnov (KS) test, reporting both the test statistic D and its p-value. To summarize the results across features, D values for the proximal 100 bp region are displayed in Supplementary Fig. 11, sorted from largest to smallest.

Generative AI and AI-assisted technologies

During the preparation of this work, the authors used Claude and ChatGPT to improve language and readability. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Reporting summary

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

Data availability

All the data generated in this study are available within the paper, its Supplementary Information, and the Source Data file provided. Source data are provided with this paper.

Code availability

Custom MATLAB scripts used for data analysis are available at DOI: 10.5281/zenodo. 17474620.

References

1. Neph, S. et al. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83–90 (2012).
2. Wang, J. et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* **22**, 1798–1812 (2012).
3. Brodsky, S. et al. Intrinsically disordered regions direct transcription factor in vivo binding specificity. *Mol. Cell* **79**, 459–471.e4 (2020).
4. Brodsky, S., Jana, T. & Barkai, N. Order through disorder: the role of intrinsically disordered regions in transcription factor binding specificity. *Curr. Opin. Struct. Biol.* **71**, 110–115 (2021).
5. Kumar, D. K. et al. Complementary strategies for directing in vivo transcription factor binding through DNA binding domains and intrinsically disordered regions. *Mol. Cell* **83**, 1462–1473.e5 (2023).
6. Boija, A. et al. Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* **175**, 1842–1855.e16 (2018).
7. Staller, M. V. et al. Directed mutational scanning reveals a balance between acidic and hydrophobic residues in strong human activation domains. *Cell Syst.* **13**, 334–345.e5 (2022).
8. Garcia, D. A. et al. An intrinsically disordered region-mediated confinement state contributes to the dynamics and function of transcription factors. *Mol. Cell* **81**, 1484–1498.e6 (2021).
9. Baughman, H. E. R. et al. An intrinsically disordered transcription activation domain increases the DNA binding affinity and reduces the specificity of NFκB p50/RelA. *J. Biol. Chem.* **298**, 102349 (2022).
10. Krois, A. S., Park, S., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Mapping interactions of the intrinsically disordered C-terminal regions of tetrameric p53 by segmental isotope labeling and NMR. *Biochemistry* **61**, 2709–2719 (2022).
11. Suter, D. M. Transcription factors and DNA play hide and seek. *Trends Cell Biol.* **30**, 491–500 (2020).
12. Mirny, L. et al. How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A Math. Theor.* **42**, 434013 (2009).
13. Berg, O. G., Winter, R. B. & von Hippel, P. H. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. models and theory. *Biochemistry* **20**, 6929–6948 (1981).
14. Garcia, D. A. et al. Power-law behavior of transcription factor dynamics at the single-molecule level implies a continuum affinity model. *Nucleic Acids Res.* **49**, 6605–6620 (2021).
15. Pomp, W., Meeussen, J. V. W. & Lenstra, T. L. Transcription factor exchange enables prolonged transcriptional bursts. *Mol. Cell* **84**, 1036–1048.e9 (2024).
16. Normanno, D. et al. Probing the target search of DNA-binding proteins in mammalian cells using TetR as a model searcher. *Nat. Commun.* **6**, 7357 (2015).
17. Zandarashvili, L. et al. Asymmetrical roles of zinc fingers in the dynamic DNA-scanning process by the inducible transcription factor Egr-1. *Proc. Natl. Acad. Sci. USA.* **109**, E1724–E1732 (2012).
18. Marklund, E. et al. DNA surface exploration and operator bypassing during target search. *Nature* **583**, 858–861 (2020).
19. Marklund, E. et al. Sequence specificity in DNA binding is mainly governed by association. *Science* **375**, 442–445 (2022).
20. Jonas, F., Navon, Y. & Barkai, N. Intrinsically disordered regions as facilitators of the transcription factor target search. *Nat. Rev. Genet.* **26**, 424–435 (2025).
21. Wagh, K., Stavreva, D. A., Upadhyaya, A. & Hager, G. L. Transcription factor dynamics: one molecule at a time. *Annu. Rev. Cell Dev. Biol.* **39**, 277–305 (2023).
22. Dror, I., Golan, T., Levy, C., Rohs, R. & Mandel-Gutfreund, Y. A widespread role of the motif environment in transcription factor binding across diverse protein families. *Genome Res.* **25**, 1268–1280 (2015).
23. Afek, A., Schipper, J. L., Horton, J., Gordân, R. & Lukatsky, D. B. Protein-DNA binding in the absence of specific base-pair recognition. *Proc. Natl. Acad. Sci. USA.* **111**, 17140–17145 (2014).
24. Levo, M. et al. Unraveling determinants of transcription factor binding outside the core binding site. *Genome Res.* **25**, 1018–1029 (2015).
25. Gasch, A. P. et al. Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257 (2000).
26. Erdős, G., Pajkos, M. & Dosztányi, Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res.* **49**, W297–W303 (2021).
27. Rudnizky, S. et al. Single-molecule DNA unzipping reveals asymmetric modulation of a transcription factor by its binding site sequence and context. *Nucleic Acids Res.* **46**, 1513–1524 (2018).
28. Khamis, H., Rudnizky, S., Melamed, P. & Kaplan, A. Single molecule characterization of the binding kinetics of a transcription factor and its modulation by DNA sequence and methylation. *Nucleic Acids Res.* **49**, 10975–10987 (2021).
29. Chen, W., Lu, W., Wolynes, P. G. & Komives, E. A. Single-molecule conformational dynamics of a transcription factor reveals a continuum of binding modes controlling association and dissociation. *Nucleic Acids Res.* **49**, 11211–11223 (2021).

30. Berg, O. G. & von Hippel, P. H. Diffusion-controlled macromolecular interactions. *Annu. Rev. Biophys. Biophys. Chem.* **14**, 131–158 (1985).
31. Slutsky, M. & Mirny, L. A. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* **87**, 4021–4035 (2004).
32. Hammar, P. et al. The lac repressor displays facilitated diffusion in living cells. *Science* **336**, 1595–1598 (2012).
33. Elf, J., Li, G.-W. & Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194 (2007).
34. Feng, X. A. et al. GAGA zinc finger transcription factor searches chromatin by 1D–3D facilitated diffusion. *Nat. Struct. Mol. Biol.* **32**, 2359–2370 (2025).
35. Ji, W., Hachmo, O., Barkai, N. & Amir, A. Design principles of transcription factors with intrinsically disordered regions. *eLife* **14**, RP104956 (2025).
36. Misiura, M. M. & Kolomeisky, A. B. Role of intrinsically disordered regions in acceleration of protein–protein association. *J. Phys. Chem. B* **124**, 20–27 (2020).
37. Arbesú, M., Iruela, G., Fuentes, H., Teixeira, J. M. C. & Pons, M. Intramolecular fuzzy interactions involving intrinsically disordered domains. *Front. Mol. Biosci.* **5**, 39 (2018).
38. Gao, A. et al. Evolution of weak cooperative interactions for biological specificity. *Proc. Natl. Acad. Sci. USA.* **115**, E11053–E11060 (2018).
39. Jonas, F. et al. The molecular grammar of protein disorder guiding genome-binding locations. *Nucleic Acids Res.* **51**, 4831–4844 (2023).
40. Wu, T., Lyu, R., You, Q. & He, C. Kethoxal-assisted single-stranded DNA sequencing captures global transcription dynamics and enhancer activity in situ. *Nat. Methods* **17**, 515–523 (2020).
41. Crossley, M. P., Bocek, M. & Cimprich, K. A. R-loops as cellular regulators and genomic threats. *Mol. Cell* **73**, 398–411 (2019).
42. Leicher, R. et al. Single-stranded nucleic acid binding and coacervation by linker histone H1. *Nat. Struct. Mol. Biol.* **29**, 463–471 (2022).
43. Wilson, K. A., Holland, D. J. & Wetmore, S. D. Topology of RNA–protein nucleobase–amino acid π–π interactions and comparison to analogous DNA–protein π–π contacts. *RNA* **22**, 696–708 (2016).
44. Rudnizky, S. et al. H2A.Z controls the stability and mobility of nucleosomes to regulate expression of the LH genes. *Nat. Commun.* **7**, 12958 (2016).
45. Rudnizky, S., Khamis, H., Malik, O., Melamed, P. & Kaplan, A. The base pair-scale diffusion of nucleosomes modulates the binding of transcription factors. *Proc. Natl. Acad. Sci. USA.* **116**, 12161–12166 (2019).
46. Rudnizky, S. et al. Extended and dynamic linker histone-DNA interactions control chromatosome compaction. *Mol. Cell* **81**, 3410–3421.e4 (2021).
47. Nguyen, T. et al. Chromatin sequesters pioneer transcription factor Sox2 from exerting force on DNA. *Nat. Commun.* **13**, 3988 (2022).
48. Ferrie, J. J., Karr, J. P., Tjian, R. & Darzacq, X. Structure"-function relationships in eukaryotic transcription factors: the role of intrinsically disordered regions in gene regulation. *Mol. Cell* **82**, 3970–3984 (2022).
49. Echave, J., Spielman, S. J. & Wilke, C. O. Causes of evolutionary rate variation among protein sites. *Nat. Rev. Genet.* **17**, 109–121 (2016).
50. Malik, O., Khamis, H., Rudnizky, S., Marx, A. & Kaplan, A. Pausing kinetics dominate strand-displacement polymerization by reverse transcriptase. *Nucleic Acids Res.* **45**, 10190–10205 (2017).
51. Gittes, F. & Schmidt, C. F. Interference model for back-focal-plane displacement detection in optical tweezers. *Opt. Lett.* **23**, 7–9 (1998).
52. Tolić-Nørrelykke, S. F. et al. Calibration of optical tweezers with positional detection in the back focal plane. *Rev. Sci. Instrum.* **77**, 103101 (2006).
53. Brouwer, I. et al. Probing DNA–DNA interactions with a combination of quadruple-trap optical tweezers and microfluidics. In *Optical Tweezers. Methods in Molecular Biology* (Gennerich, A. eds) Vol. 1486, 275–293 (Humana Press, New York, NY, 2017) https://doi.org/10.1007/978-1-4939-6421-5_10.
54. Marko, J. F. & Siggia, E. D. Stretching DNA. *Macromolecules* **28**, 8759–8770 (1995).
55. Agresti, A. & Coull, B. A. Approximate is better than 'exact' for interval estimation of binomial proportions. *Am. Stat.* **52**, 119–126 (1998).
56. Bronson, J. E., Fei, J., Hofman, J. M., Gonzalez, R. L. & Wiggins, C. H. Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.* **97**, 3196–3205 (2009).
57. McKinney, S. A., Joo, C. & Ha, T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **91**, 1941–1951 (2006).
58. Fornes, O. et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).
59. Li, J. & Rohs, R. Deep DNAshape webserver: prediction and real-time visualization of DNA shape considering extended k-mers. *Nucleic Acids Res.* **52**, W7–W12 (2024).
60. Li, J., Chiu, T.-P. & Rohs, R. Predicting DNA structure using a deep learning method. *Nat. Commun.* **15**, 1243 (2024).

Acknowledgements

We thank Hagen Hofmann, Naama Barkai, and Arnon Henn for valuable comments, and Gabriel Rosenblum, Vladimir Mindel and Moshe Goldsmith for their assistance. This work was supported by the Israel Science Foundation (Grants 937/20 and 255/25 to AK).

Author contributions

A.K. and N.S. conceived the project. N.S. and C.B. designed and performed all experiments. N.S., C.B., H.K., and A.K. analyzed the data. N.S., C.B., S.S.H., and N.N. prepared experimental materials. N.S., C.B., and A.K. prepared the figures. A.K. and N.S. wrote the manuscript. A.K. supervised the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-67217-2.

Correspondence and requests for materials should be addressed to Ariel Kaplan.

Peer review information *Nature Communications* thanks Shixin Liu, Anatoly Kolomeisky, Katherine Stott, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available..