PCCP

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Phys. Chem. Chem. Phys.*, 2025, **27**, 26107

Energetic and structural dynamic drivers of transcription factor MycMax, Omomyc homodimer, and MaxMax recognition on DNA

Carmen Al Masri[a] and Jin Yu*[b]

Basic helix–loop–helix leucine zipper (bHLHLZ) transcription factors (TFs), such as MycMax, MaxMax, and the engineered mutant Omomyc dimers, regulate gene expression by binding to the E-box (5′-CACGTG-3′) motif in DNA. Here, we used microsecond molecular dynamics (MD) simulations along with MMGBSA energetic and hydrogen bonding (HB) analyses to investigate how MycMax, Omomyc, and MaxMax dimers interact with both canonical E-box and homogeneous polyA reference sequences. Our MD results indicate that protein–DNA van der Waals (VDW) interactions and surface complementarity dominate the binding affinity measurements. Per-residue decomposition underscores the central role of conserved arginine residues at the protein–DNA binding interface, while HB information readout (or KL divergence) analyses reveal a biased DNA strand association that reinforces protein sequence recognition for specificity. Although MaxMax binds to the canonical E-box energetically stably, it loses the stability on polyA DNA more than the other two TFs, suggesting an essential reliance of MaxMax on the E-box motif for TF–DNA interaction coordination. MycMax exhibits moderate stability energetically and shows more variable dynamical responses to local DNA sequences. In contrast, the Omomyc dimer retains slightly higher binding affinity or energetic stability than MycMax on both E-box and polyA DNA, suggesting an enhanced ability to sequester MycMax in oncogenic settings. These findings provide new physical insights into dimeric bHLHLZ–TFs on DNA recognition and reveal potential strategies for targeting Myc-driven transcription or overexpression in cancer.

Received 19th September 2025,
Accepted 8th November 2025

DOI: 10.1039/d5cp03629a

rsc.li/pccp

Introduction

Basic helix–loop–helix leucine zipper (bHLHLZ) transcription factors (TFs) (Fig. 1A) govern key cellular processes such as cell proliferation, differentiation, and apoptosis by binding to specific DNA sequences.[1,2] Dysregulation of bHLHLZ proteins is implicated in numerous diseases; for example, overexpression of the Myc family proteins is observed in approximately 70% of human tumors and is a key driver of oncogenesis.[3–5] Myc heterodimerizes with Max *via* their bHLHLZ domains to bind to a canonical E-box motif (5′-CACGTG-3′) in gene promoters, driving a broad transcriptional program that influences cell growth, metabolism, and the cell cycle.[2,4,6] In contrast, Max can homodimerize and repress transcription by competing for the same E-box sites, thus balancing Myc's oncogenic effects.[1,7] Other bHLHLZ proteins, such as the Mxd family, further modulate Myc activity by forming repressive complexes.[1,2]

A promising strategy to inhibit Myc in cancer is *via* Omomyc, a synthetic 91-amino-acid protein mutant derived from Myc's bHLHLZ domain.[8] Omomyc was engineered by substituting four key charged residues (two Asp and two Arg) in the leucine zipper (LZ), which normally prevent Myc homodimerization, with those found mainly in Max[8] (see the mutated residues T, I, Q, and N in Omomyc in Fig. 1B and C). This design enables Omomyc to efficiently homodimerize as well as to heterodimerize with both Myc and Max.[9–11] As a result, Omomyc can compete with or prevent Myc from both dimerizing with Max and binding to E-box sequences, thereby reducing or disrupting Myc-driven gene expression. Preclinical studies have shown that Omomyc causes significant tumor regression with minimal and reversible side effects on normal tissues,[5,11] suggesting its potential as an effective Myc inhibitor. Notably, Omomyc has advanced to clinical evaluation; OMO-103, the clinical formulation of Omomyc, successfully completed a phase I trial in patients with advanced solid tumors, demonstrating safety and preliminary signs of anti-tumor activity.[12]

---

[a] *Department of Physics and Astronomy, University of California, Irvine, California, USA*

[b] *Department of Physics and Astronomy, Department of Chemistry, NSF-Simons Center for Multiscale Cell Fate Research, University of California, Irvine, California, USA. E-mail: jin.yu@uci.edu*

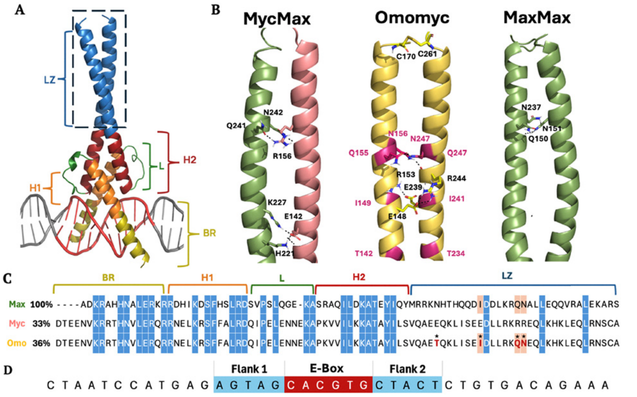

Fig. 1 Structural and sequence comparison of MycMax, Omomyc dimer, and MaxMax proteins. The structures were obtained for MycMax (PDB: 1NKP[19]), the Omomyc dimer (PDB: 5I50[9]), and MaxMax (PDB: 1AN2[21]) from the Protein Data Bank (PDB). (A) Schematic representation of the DNA-bound MycMax bHLHLZ domain, highlighting the basic region (BR, yellow), helix 1 (H1, orange), loop (L, green), helix 2 (H2, red), and leucine zipper (LZ, blue). The E-box on the DNA is colored in red. (B) Molecular view of the HB interactions within the LZ regions of the MycMax (left), Omomyc dimer (center), and MaxMax (right) complexes. The four mutations (N156/N247, Q155/Q247, I149/I241, and T142/T234) in Omomyc with respect to Myc are highlighted in magenta. (C) Sequence alignment of the bHLHLZ domains from Max, Myc, and Omomyc. Conserved residues are highlighted in blue. The four mutated residues in the Omomyc LZ are marked with a star and colored in red. Omomyc residues from mutations that align with Max are highlighted in pink. The percentages indicate Myc and Omomyc sequence identity compared to Max. (D) DNA sequence utilized in modelling protein binding, including the central core (E-box, highlighted in red) and flanking sequence (highlighted in blue) regions.

Although Myc, Max, and Omomyc share a common bHLHLZ architecture, differences in their LZ and basic DNA-binding regions (BRs) critically influence their dimerization strength and DNA-binding specificity, respectively. For example, the four point mutations (E142T, E149I, R155Q, and R156N) in Omomyc not only promote homodimerization by mimicking key Max residues but also enhance dimerization stability through additional hydrophobic and hydrogen-bond (HB) interactions (Fig. 1B, C and Fig. S1).[8,9] However, while crystallographic and biochemical studies have detailed the static interactions, the dynamic and energetic determinants that govern how these dimers selectively recognize specific (E-box) *versus* nonspecific DNA remain incompletely understood. It is also unclear how bHLHLZ TF–DNA interfacial contacts modulate binding affinity or specificity in different DNA contexts and which of the dimers is the most stabilized energetically bound on E-box or nonspecific DNA.

To address these questions, we performed microsecond all-atom MD simulations of MycMax, MaxMax, and Omomyc dimers in complex with both a canonical E-box (CACGTG) and a homogeneous polyadenine (polyA) 40-base pair (bp) sequence. PolyA was chosen as a control or reference sequence with minimal information entropy, and it has also been confirmed from protein binding microarray (PBM) experiments that polyA is among the most weakly bound by bHLHLZ proteins.[13] Additionally, its relatively uniform DNA shape removes bias from sequence-dependent shape recognition, making it a useful control highlighting nonspecific interactions or comparing with specific interactions. By analyzing protein–DNA conformational dynamics, interfacial contacts and HB patterns, and binding free energies or components computed *via* the MMGBSA method,[14] we characterized how each dimer adapts to E-box *versus* polyA DNA context. Our simulations reveal that although MaxMax binds the most stably energetically on E-box DNA, it undergoes the most energetic destabilization on polyA, whereas Omomyc and MycMax maintain their relative stabilities on DNA, with Omomyc slightly more stabilized than MycMax no matter on E-box or polyA. These observations can be compatible with early experimental findings on relatively high DNA binding affinity (with a low dissociation constant) of Max monomer and dimer to the E-box[15] as well as later studies showing overall low DNA binding stability of MaxMax, likely due to its notable dimerization instabilities.[9,15,16] Additionally, it has been found that MycMax binds to weaker motifs at high concentrations,[17,18] and Omomyc effectively occupies low-affinity promoter regions, thereby displacing Myc from DNA.[10]

Overall, our findings provide new physical insights into how the protein sequence variations within the bHLHLZ BR or DNA binding domain (DBD) influence DNA binding stability and specificity, deepen our understanding of Myc-mediated transcriptional regulation, and offer a framework for improving rational design of next-generation therapeutics to target Myc-driven tumors.

Materials and methods

System setup. The structures for MycMax (PDB: 1NKP[19,20]), the Omomyc dimer (PDB: 5I50[9]), and MaxMax (PDB: 1AN2[21]) bound to E-box DNA were obtained from the Protein Data Bank (PDB). To ensure consistency with experimentally expressed protein sequences, we aligned the bHLHLZ domains from each PDB structure with corresponding UniProt[22] entries (P01106 for Myc and P61244 for Max), using PyMOL[23] to substitute or extend any missing or engineered residues as needed. A protein residue alignment between the PDB, UniProt, and the current structures, demonstrating which residues were deleted or inserted, is available in Fig. S2, and a residue number mapping in Fig. S3. To mitigate DNA end effects, each system's DNA was extended to 40 bps with a random sequence using the composite module of web 3DNA.[24] For polyA systems, all nucleotides in the 40-bp DNA were mutated to adenine using the mutation module of web 3DNA. Protonation states were assigned using pdb2pqr[25] at a physiological pH 7.4. All the processed system structural PDB files are available in the SI.

MD simulations. The E-box-bound systems were simulated using GROMACS 2022.1,[26,27] and the polyA-bound systems were simulated on Anton2 supercomputer for a longer timescale.[28] For the purpose of the current work, we conducted and analyzed one-microsecond simulation for each system. All simulations employed the AMBER14SB[29] force field for proteins and the OL15[30] force field for DNA, with water modeled using TIP3P.

All systems underwent initial processing, minimization, and equilibration using GROMACS. Each system was placed in a periodic box with a minimum distance of 1.5 nm to the box edge, solvated, and ionized with Na^+ and Cl^- ions to a final ionic strength of 0.15 M. A rectangular box was used for E-box simulations, and a cubic box was used for the polyA simulations to be compatible with Anton2. The total number of atoms per system was approximately 140 000 for E-box systems and about 490 000 for polyA systems. Short-range electrostatic and van der Waals (VDW) interactions were treated with a 1.2 nm cutoff, while long-range electrostatics were computed using the particle-mesh Ewald[31] (PME) method.

Energy minimization was performed using the steepest-descent algorithm. The systems were then equilibrated in two stages: first for 4 ns under the NVT ensemble, followed by another 4 ns under the NPT ensemble, using a time step of 2 fs. Heavy-atom position restraints were applied with a force constant of 1000 kJ mol^{-1} nm^{-2}. Temperature was maintained at 300 K using the leap-frog stochastic dynamics integrator with an inverse friction constant of 2 ps, while pressure was controlled at 1 bar using the Parrinello–Rahman barostat.[32]

For E-box systems, production simulations were performed for 1 μs in GROMACS, applying restraints only to the first and last heavy atoms of DNA to maintain alignment along the *x*-axis. For polyA systems, an additional 30 ns equilibration with DNA end restraints was carried out in GROMACS, after which the systems were transferred to Anton2 for a 10-μs production run. Only the first 1 μs of this trajectory is reported in the current study.

MMGBSA calculations. As a computational measure of the protein–DNA binding affinity, the binding free energies (ΔG_{bind}) were estimated using the molecular mechanics/generalized born surface area (MMGBSA)[14,33] approach as implemented in MMGBSA.py[34] *via* the gmx_MMPBSA tool.[35] In the calculations, the protein–DNA binding free energy in the solvated state, ΔG_{bind} (<0), is estimated as a linear summation of individual energetics:

$$\Delta G_{\mathrm{bind}} = \Delta G_{\mathrm{bind,vacuum}} + \Delta G^{C}_{\mathrm{solvation}} - (\Delta G^{P}_{\mathrm{solvation}} + \Delta G^{D}_{\mathrm{solvation}}) \tag{1}$$

The binding free energy in vacuum $\Delta G_{\mathrm{bind,vacuum}}$ is decomposed into the gas-phase molecular mechanics energy ΔE_{MM} and the configurational entropy part of the protein–DNA system $T\Delta S$:

$$\Delta G_{\mathrm{bind,vacuum}} = \Delta E_{\mathrm{MM}} - T\Delta S \tag{2}$$

ΔE_{MM} is taken as the sum of the VDW (ΔE_{VDW}), electrostatic (ΔE_{EES}) and bonded ($\Delta E_{\mathrm{bonded}}$) interaction energies calculated in the absence of solvent. Owing to the single-trajectory approach (*i.e.*, both bound and unbound states are taken from the same trajectory), the protein–DNA complex does not undergo a change in conformational strain, which leads to $\Delta E_{\mathrm{bonded}} = 0$. The electrostatic term quantifies the direct Coulombic interactions between atomic partial charges, while the VDW term accounts for dispersion forces and steric repulsion. Both terms are calculated from the current force field.

Entropic contributions were incorporated using the interaction entropy method,[36] which estimates the relative entropy or entropy change before and after protein–ligand (or DNA nucleotides here) binding $-T\Delta S$ from fluctuations in the gas-phase protein–DNA interactions (ΔE_{MM}) over the trajectory as $k_{\mathrm{B}}T\exp\left(\dfrac{\Delta E_{\mathrm{MM}} - \Delta E_{\mathrm{MM}}}{k_{\mathrm{B}}T}\right)$ with $\langle\ldots\rangle$ denoting an ensemble average. In the calculations, the protein–DNA interaction energy ΔE_{MM} is regarded as the difference of gas-phase energy between that of the protein–DNA complex and those of the separate protein and DNA, which can be directly evaluated in the simulation of the protein–DNA complex. This interaction entropy method offers an efficient alternative to conventional normal mode analysis, providing improved accuracy in capturing interaction conformational entropy effects.[37]

The solvation free energy, $\Delta G_{\mathrm{solvation}}$, is partitioned into polar and nonpolar components for the complex (C), the protein (P) and the DNA (D):

$$\Delta G^{\mathrm{C/P/D}}_{\mathrm{solvation}} = \Delta G^{\mathrm{C/P/D}}_{\mathrm{polar}} + \Delta G^{\mathrm{C/P/D}}_{\mathrm{nonpolar}} \tag{3}$$

The polar solvation free energy contribution $\Delta G_{\mathrm{polar}}$ is estimated using the Generalized Born (GB) approximation.[38,39] This model treats the solvent as a continuum dielectric, computing the electrostatic solvation energy based on effective Born radii and interatomic distances to quantify the degree of burial of individual atoms. The nonpolar solvation free energy contribution $\Delta G_{\mathrm{nonpolar}}$ is calculated from a linear relationship with the solvent-accessible surface area (SASA), scaled by a surface tension parameter.[35] Altogether, the larger the magnitude of the total free energetics $|\Delta G_{\mathrm{bind}}|$ (see eqn (1)), the larger the protein–DNA binding affinity. Additionally, the binding specificity can be measured by the relative protein binding free energy between the DNA of concern and a reference DNA (*i.e.*, $\Delta\Delta G_{\mathrm{bind}} = \Delta G_{\mathrm{bind}} - \Delta G^{\mathrm{ref}}_{\mathrm{bind}}$). However, note that the MMGBSA approach employs continuum solvation model approximation and assumes linear superposition of individual energetics (eqn (1)); limited computational sampling also prevents accurate entropy calculation. Hence, MMGBSA often does not provide highly accurate prediction of the binding free energy.[40]

All MMGBSA calculations were performed on the entirety of 1-μs production runs, with one frame sampled every nanosecond (1000 frames in total). To appropriately capture the high charge density effect inherent in DNA, the internal dielectric constant *intdiel* was set to 10 while the external dielectric was fixed at 78.5. The choice of internal dielectric was based on our benchmarks (Fig. S4), which showed that the total binding free energy convergence was achieved starting *intdiel* = 10.

Results and discussion

Structural dynamics stability of dimeric TFs on E-box and polyA DNA

We first examined how MycMax, Omomyc, and MaxMax bHLHLZ dimers adapt to binding either a canonical E-box (5′-CACGTG-3′) or a polyA (5′-AAAAAA-3′) DNA sequence at

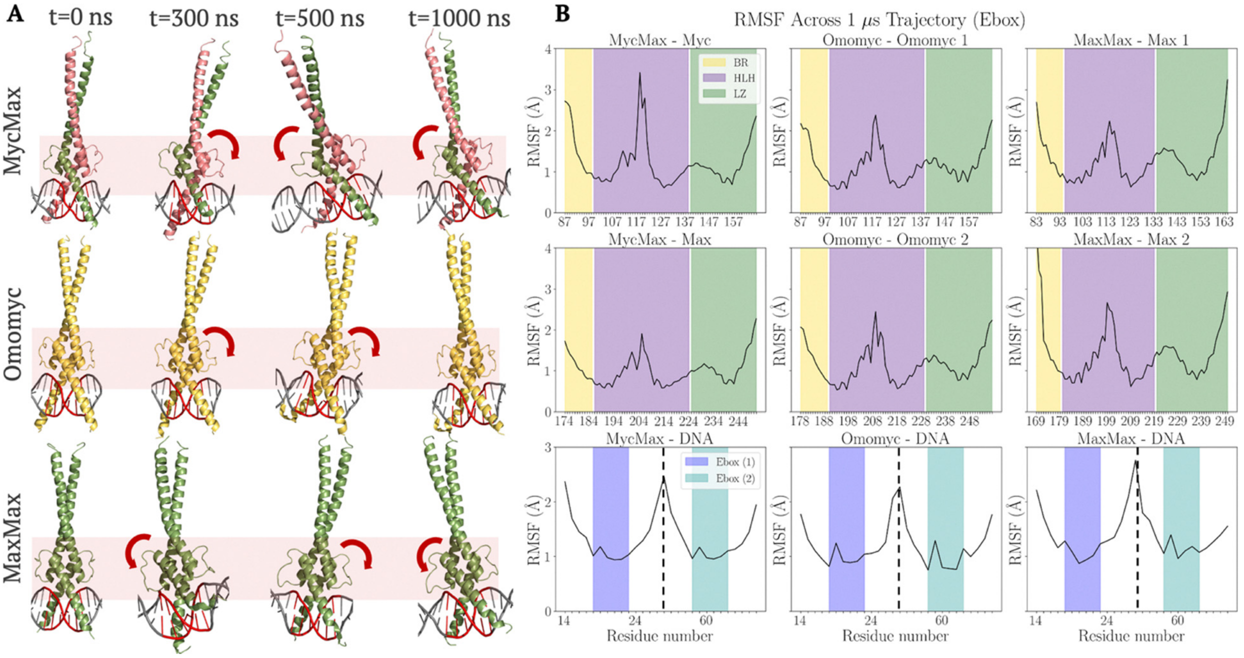

Fig. 2 Equilibrium structural dynamics and flexibility of the MycMax, Omomyc dimer, and MaxMax complexes. (A) Snapshots of the E-box-bound MycMax, Omomyc, and MaxMax complexes captured at different simulation times (t = 0 ns, 300 ns, 500 ns, and 1000 ns). Myc is colored in pink, Max in green and Omomyc in yellow. The E-box is shown in red. The HLH region is highlighted by the red band. Red arrows illustrate the rocking motion of the dimers. (B) RMSF plots for MycMax, Omomyc, and MaxMax–DNA complexes. RMSF is measured for each protein monomer (top two rows), as well as for both DNA strands (bottom row, with two strands indexed by nucleotide residue numbers and delimited by dashed lines), over the entire 1 μs simulation period. The protein regions are color-coded: BR (yellow), HLH (purple), LZ (green), and the E-box on DNA with two single-stranded (ss) oligos 1 (purple, labeled Ebox (1)) and 2 (blue, labelled Ebox (2)).

the center of a 40-bp length DNA construct. Despite sharing the same core bHLHLZ fold, the three complexes displayed both common and distinct structural dynamics over the microsecond-scale simulations. Note that the structural dynamics stability can be directly monitored by root-mean square deviation (RMSD) of protein and/or DNA backbone structures (shown in Fig. S5).

MycMax exhibits the pronounced HLHLZ rocking motion on DNA. In E-box–bound systems, one monomer's BR in all three dimers deviated markedly at around 400 ns (Fig. S5), increasing RMSD by ∼2 Å and coinciding with transient contacts to flanking DNA segments (see also HB Patterns and KL Divergence). This deviation correlated with a "rocking motion" of the HLHLZ domains along the DNA axis, wherein the HLH portion periodically shifted to engage on DNA flanking sequences (Fig. 2A and Fig. S6, S7). Notably, the amplitude of this rocking motion varied with the TF–DNA complex (Fig. S7), with MycMax showing the largest angular deviation (16 ± 9°), followed by MaxMax (12 ± 5°) and Omomyc (11 ± 6°). Upon switching to polyA, the rocking motion diminished significantly for Omomyc and MaxMax (7 ± 4° and 5 ± 3°, respectively) yet remained highly pronounced for MycMax (14 ± 3°). Trajectories are provided in the SI to visualize such motions.

Note that there is a lysine residue (K57) on a loop from HLH identified in the MaxMax crystal structure (PDB: 1AN2[21]) contacting the DNA backbone in the minor groove that is also involved in the rocking motion, which corresponds to K207 in Max and K121 in Myc of MycMax modelled, or K121 and K212 in the Omomyc dimer, and K116 and K202 in MaxMax (see numbering convention in Fig. S3).

MycMax exhibits large Myc–BR structural deviations or destabilizations on polyA–DNA. All three complexes maintained relatively stable HLH and LZ domain motions (from Fig. S5: RMSD = 2–3 Å for HLH, 3–4 Å for LZ). In contrast, the BRs showed large deviations, especially on polyA. Notably, MycMax on polyA exhibited the largest deviation in Myc–BR (∼8.6 Å), substantially exceeding that of Omomyc (∼5 Å) or MaxMax (∼4 Å) on polyA (Fig. S5B). These elevated deviations in MycMax also propagated to the DNA: MycMax–polyA complexes had a higher DNA RMSD (∼4–5 Å) than other DNA counterparts (∼2.5 Å). By contrast, Omomyc and MaxMax induced only small perturbations in polyA DNA. The two BRs in each dimeric TF also displayed disparate RMSDs except for the MaxMax–polyA system (Fig. S5B).

MaxMax LZs exhibit high and low structural flexibility respectively on E-box and polyA. In addition, we used root-mean-square fluctuation (RMSF), *i.e.*, by averaging atomic coordinate fluctuations of individual residues (amino acids or nucleotides) over MD trajectories to reveal local structural flexibility (Fig. 2B). As the overall RMSFs started converging after ∼300 ns and completely converged by 700 ns (Fig. S8), the magnitudes of RMSFs differed notably: on the E-box, MaxMax LZs displayed the highest RMSFs (∼3–4 Å), while the LZs on MycMax and Omomyc were fairly constrained (∼2 Å). This trend changed on polyA, with MaxMax's LZ along with HLH RMSFs decreasing and Omomyc's rising to ∼4 Å (Fig. S9).

MycMax HLHs and BRs show asymmetry on RMSFs between monomers. Consistently, homodimers showed rather symmetrical RMSF profiles, whereas heterodimer MycMax on E-box exhibited asymmetry between two monomers, especially in its HLH region (Fig. 2B, top two rows left). Binding to the E-box also dampened fluctuations in the DNA's central 6-bp across all three complexes (Fig. 2B, bottom row), stabilizing the duplex relative to unbound DNA. While slightly weaker, the central 6-bp stabilizing effect persists on polyA (Fig. S9), whereas the Myc–BR flexibility significantly increased by ∼1–2 Å and HLH RMSF peaks slightly decreased on polyA (Fig. S9).

Overall, the structural dynamics along with stability and flexibilities of three complexes underscore how the dimeric bHLHLZ domain architectures are modulated by protein–DNA interactions. In particular, MycMax as a heterodimer displays notable asymmetry of flexibilities between its HLH domains on the E-box and very large deviations of Myc–BR occasionally on poly-A, while overall it shows significant LZ rocking motions on both DNA types. The pronounced rocking motion of MycMax on the E-box and occasional large BR deviation on polyA together indicate some structural instabilities and consequent energetic instabilities of MycMax on DNA, as we analysed below (by ΔG_{bind}). Note that current simulations were conducted with only the ordered structural region of Myc modeled. The structural dynamics properties may however be modified if Myc becomes full-length including the disordered region. Additionally, we observed that the LZ region of the MaxMax homodimer exhibits greater flexibility on E-box DNA compared to poly-A DNA. This increased LZ flexibility may correlate with increased TF–DNA binding affinity, as MaxMax is found energetically more stable (by ΔG_{bind}) on E-box DNA compared to MycMax and Omomyc, with the trend reversed on poly-A DNA (see the next subsection). These results thus set the stage for further investigations into the energetic basis of TF–DNA binding and the particular roles of key residues and flanking sequences in modulating the binding affinity and specificity.

TF–DNA binding energetic stability analysis *via* MMGBSA

MMGBSA provides a fast method to estimate binding free energies, although it is also known to overestimate the energies[33] and yield large standard deviations (in our case, on the order of tens of K_BT for ΔG_{bind}). In the current study, we calculated TF–DNA binding free energy components *via* MMGBSA (eqn (1)) over the microsecond-long MD trajectories. We compared averages computed over the entire trajectory with those from earlier segments (100–300 ns, *i.e.*, before major conformational changes in the BR occurred) and found that the overall energy trends remained unchanged (data not shown). From Fig. S10 and S11, we can see that around 100–200 ns, the total energy trends become consistent. Consequently, we consider our MMGBSA estimates to be significantly converged and report the averages over the entire trajectory.

VDW dominance in TF–DNA binding energetics and enhanced stability of MaxMax on E-box DNA. MMGBSA calculations showed that the protein–DNA VDW interactions dominated the total binding free energy (ΔG_{bind} from eqn (1)) for MycMax, the Omomyc dimer, and MaxMax stabilities on both E-box and polyA DNA (Fig. 3A and B, insets; with all the time-dependent energetic measurements shown in Fig. S10 and S11). This is consistent with prior computational and experimental studies, which report that van der Waals and nonpolar (desolvation-driven) contributions dominate protein–DNA binding free energies.[33,41,42]

In contrast, electrostatic (EEL) and polar solvation (EGB) terms largely offset each other, resulting in small net electrostatic contributions similarly in three complexes. Consequently, the overall ranking of binding free energies closely follows the VDW strength and correlates particularly strongly with the buried surface area (BSA) at the protein–DNA interface. A correlation analysis (Fig. 3C on MycMax, and Fig. S12 on the Omomyc dimer and MaxMax) supported this observation, showing near-perfect correlation ($r \approx 0.98$) between VDW and total binding energy ΔG_{bind}, along with substantial correlation ($r \approx 0.87$–0.89) between total binding energy and nonpolar solvation (ESURF). By contrast, EEL and EGB exhibited near-complete anti-correlation. Such patterns thus persist from MycMax (Fig. 3C) to the Omomyc dimer and MaxMax (Fig. S12). Interestingly, the correlation between the electrostatic/polar solvation terms and the other energy components varies significantly between systems (with the highest overall correlations in MaxMax and the lowest in the Omomyc dimer), however, with no clear trend directly related to overall binding stability.

Sequence-dependent variability in TF–DNA binding energetics stability. Despite Myc, Max, and Omomyc sharing largely similar bHLH regions (Fig. 1C), substituting their BR-bound DNA from the E-box to polyA yielded notably different protein–DNA stability changes (Fig. 3A and B): on E-box DNA, MaxMax displayed the highest overall binding stability to DNA, energetically, followed by Omomyc, and then MycMax. On polyA, however, Omomyc became the most stable, still energetically, MycMax was intermediate, and MaxMax dropped to the weakest binder. Notably, Omomyc ΔG_{bind} was the least affected by the E-box to polyA switch (a relatively small decrease of ∼3% in magnitude), whereas MycMax and MaxMax ΔG_{bind} decreased in magnitude by ∼6% and 20%, respectively. Such changes of binding free energetics upon DNA sequence variations reflect TF binding specificities, and it appears that MaxMax is the most specific to the E-box while Omomyc the least in reference to the polyA sequence.

Protein–DNA interfacial surface complementarity governs binding energetics stability. Our MMGBSA results underscore the importance of VDW interactions and BSA in stabilizing the TF–DNA interface, supporting the idea that protein–DNA surface complementarity is central to bHLHLZ binding and recognition.[41,43] To examine whether DNA shape contributes to this complementarity, we measured the minor and major groove widths (Fig. 3D). On the E-box, all three dimers widened the minor groove by over 2 Å and slightly narrowed the major groove at the center, as expected from specific

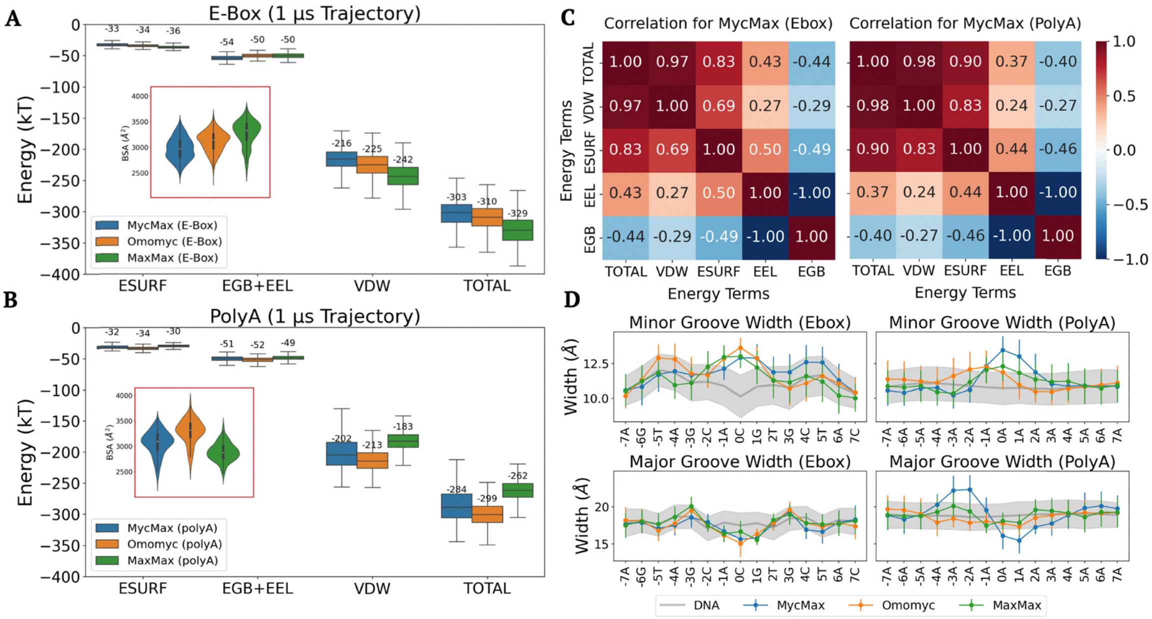

Fig. 3 TF–DNA binding (free) energetics and correlation analyses. (A) and (B) Box plots showing MMGBSA energy decomposition for MycMax (blue), the Omomyc dimer (orange), and MaxMax (green) bound to E-box (A) or polyA (B) across six energy terms: ESURF (non-polar solvation energy), EEL (electrostatic energy), EGB (polar solvation energy), EGB + EEL (sum of polar solvation and electrostatic energies), VDW (van der Waals energy), and TOTAL (total binding energy ΔG_{bind} from eqn (1)). Insets on the top right provide a closer view of VDW and TOTAL energy terms (with values labelled) for each complex. The violin plot insets represent the buried surface areas (BSA) between each TF dimer and either E-box or polyA DNA. (C) Cross-correlations between MMGBSA energy terms for MycMax on E-box (left) and polyA (right). (D) Minor (top) and major (bottom) groove widths measured along DNA position in the MycMax–DNA complex, with error bars showing the standard deviation across the trajectory. DNA alone width (without protein) is also shown in grey with a light grey band indicating its standard deviation. DNA numbering is re-scaled to have the central C in the E-box CACGTG as index 0.

arginine contacts.[43,44] Although MaxMax exhibits the largest BSA, its degree of major/minor groove deformation on E-box DNA is similar to that of MycMax and Omomyc (Fig. 3D, left), suggesting that stabilization is driven more by tight protein interactions or complementarity with DNA than by DNA shape or distortion alone.

On polyA DNA, where the unbound groove profile is uniform, MycMax induced the largest groove distortions (Fig. 3D, right). In contrast, Omomyc and MaxMax maintained groove widths similar to those of unbound DNA. Notably, these distortions did not correlate directly with changes in BSA or binding affinity. These observations again indicate that TF–DNA interactions not only reflect the DNA sequences but also reveal protein active actions and responses to modulate and fit to the DNA conformation. In essence, the TF–DNA interface is dynamic, with the TFs reshaping the DNA and supplying surface complementarity. This adaptive modulation at the TF–DNA interface likely contributes to the distinct binding stabilities observed among the complexes.

Conformational entropy penalizes TF–DNA binding without altering the stability ranking. We additionally calculated the interaction entropy contribution ($-T\Delta S$) to TF–DNA binding (see Methods), with average values ranging from approximately 10 to 20 k_BT (Fig. S13A). Note that ΔS is negative, meaning that conformational entropy decreases upon binding or penalizes TF–DNA binding. Although the entropy contributions for MycMax, the Omomyc dimer, and MaxMax show some overlap, statistical tests indicate that they differ significantly ($p < 0.05$ for most cases; Fig. S13B). Consistent with previous reports,[36] we observed a negative correlation between total binding free energy ΔG_{bind} and the magnitude of the entropy contribution ($-T\Delta S$); that is, as binding becomes more favorable energetically, the conformational entropy penalty tends to be smaller. It is important to note that these calculations account only for the interaction entropy of the protein and DNA, omitting solvent effects. Given that the entropic contributions are relatively small compared to the total binding energies and their fluctuations, they do not affect the overall binding free energy ranking among MycMax, the Omomyc dimer, and MaxMax. We therefore did not include the entropy term in further analyses.

Conserved arginines provide key energetic stabilization at the bHLHLZ–DNA interfaces

Conserved arginines dominate TF–DNA interactions and stabilize the bHLHLZ interface. Per-residue MMGBSA analysis revealed that the key protein residues contributing to binding energetic stability are highly conserved arginines across the three systems, indicating an evolutionarily conserved DNA binding interface (Fig. 4A). On E-box DNA, the MaxMax dimer exhibits the largest number of interface residues with strong negative energy contributions (below −7 k_BT per residue; Fig. 4A, top). Despite quantitative differences among the three dimers, conserved arginine residues consistently emerge as the top contributors (Fig. 4B and Fig. S14A). Arginines dominate

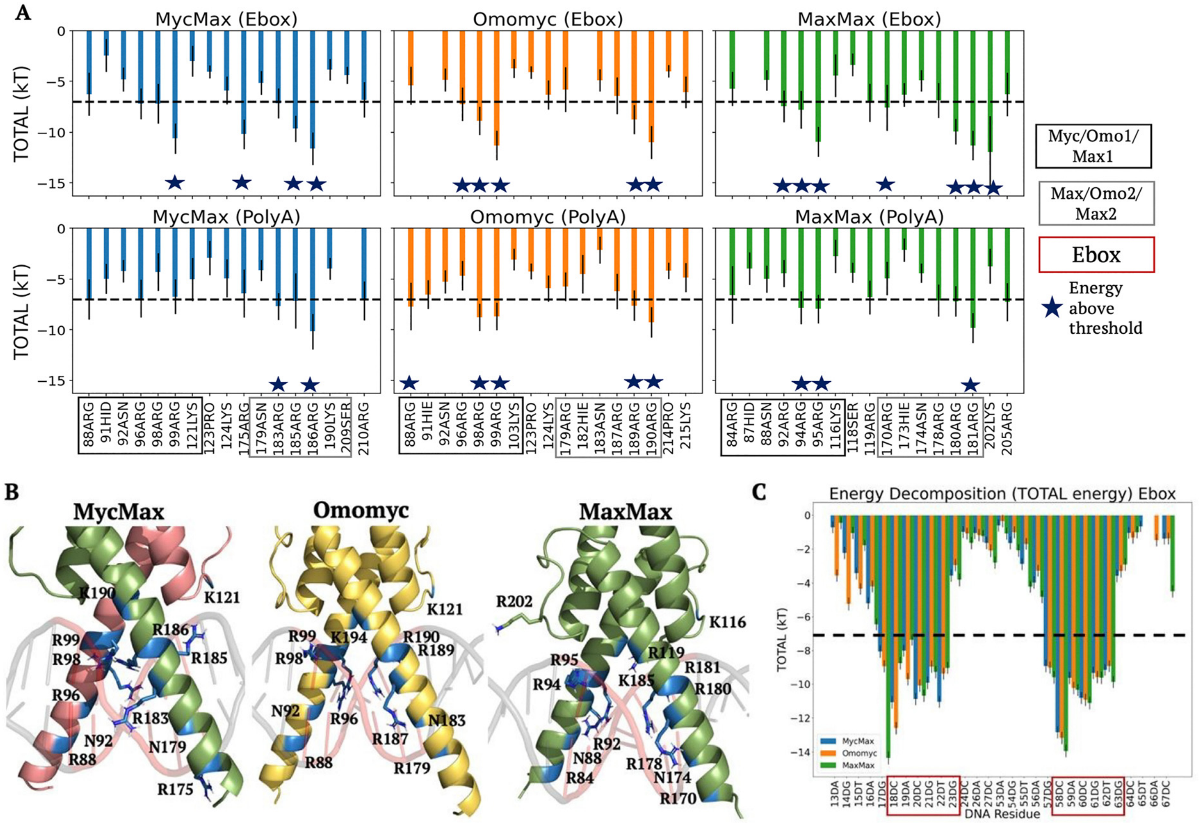

Fig. 4 Residue-level energy decomposition and key interactions in MycMax, Omomyc, and MaxMax complexes bound to E-box DNA. (A) Residue-level energy decomposition of MycMax (blue), Omomyc (orange), and MaxMax (green) bound to E-box (top) or polyA (bottom), showing the contribution of each protein residue to the total binding energy (top 15 contributing residues displayed). Key residues with energy contributions above a 7 k_BT energy threshold are marked with blue stars. The horizontal dashed line indicates the energy threshold. Residues that are conserved across all three systems are highlighted by a black box for the first monomer and a grey box for the second monomer. (B) Detailed views of the key protein–DNA interactions within the MycMax, Omomyc, and MaxMax complexes. Residues conserved across the three systems are highlighted in blue and labeled. The residues marked by a star in (A) are shown in stick representation. (C) DNA residue-level energy decomposition across MycMax, Omomyc, and MaxMax complexes, showing the contributions of DNA residues to the total binding energy. The E-box on DNA is highlighted in the red box. The horizontal dashed line indicates the energy threshold.

the energetic stabilization of the bHLHLZ–DNA interface due to their guanidinium groups, which provide multiple HB donors that enable bidentate contacts, with a high preference to guanine;[45,46] their long side chains further allow flexible engagement with diverse DNA groove geometries. Additional residues, including two lysines, one histidine, and one asparagine, also contribute, albeit with fewer contacts (Fig. 4A). When bound to polyA DNA, the overall per-residue energy contributions decrease, and fewer residues exceed the $-7\ k_BT$ threshold (Fig. 4A, bottom). Nonetheless, the conserved arginines remain the primary stabilizing elements, as confirmed by detailed energetic decomposition (Fig. S14). In the detailed energetic decomposition, arginines show the VDW along with non-polar solvent energetic decomposition similar to the total energetic decomposition, and the VDW together with non-polar solvent energetics dominant over the ELE (electrostatic) and polar solvent contributions combined.

Interactions differ more on one BR (among Myc/Omo1/Max1) than the other on the E-box. A DNA-focused decomposition (Fig. 4C) indicates that the energy contributions from interactions involving the E-box strand engaged by one monomer's BR (Myc/Omo1/Max1) vary significantly among the complexes, particularly in the flanking sequences (DNA residues outside red-boxes on the bottom axis of Fig. 4C), where Omomyc provides the largest contributions. In contrast, the energy contributions from the other monomer's interface (Max/Omo2/Max2) are more uniform across systems. For both MycMax and Omomyc dimers, especially the Omo1 subunit forms additional stabilizing contacts with flanking bases outside the core 6-bp E-box, underscoring the role of the flanking sequences in modulating binding energetics or stability, while MaxMax's interactions are more confined to the core E-box region. These trends vary somehow on polyA, which showed greater variability in per-residue energy within the 6-bp binding site than on E-box (Fig. S14B); Myc forms more stabilizing contacts, energetically more stabilized than Max, with the flanking and central binding sites on polyA, consistently revealing more asymmetric binding patterns as a heterodimer than the homodimer Omomyc and MaxMax on polyA.

It is also noteworthy that the energy decomposition for Omo1 bound to E-box shows an intermediate pattern between Myc and Max (Fig. 4A and C). This suggests that the charged residue mutations introduced to generate Omo1 not only promote homodimerization but also allosterically influence BR–DNA interactions. Finally, examination of the electrostatic (EEL) and polar solvation (EGB) contributions (Fig. S14B) confirms that these terms largely cancel at the E-box center, reinforcing that VDW interactions and overall interface complementarity are the primary determinants of TF–DNA binding energetics related stability.

HB patterns at the TF–DNA interface and KL divergence on sequence readout

Monomer-specific DNA strand association and HB patterns on the E-box *versus* polyA. We further analyzed HBs at the TF–DNA interface to understand how MycMax, the Omomyc dimer, and MaxMax engage on DNA sequence readout, *i.e.*, by comparing E-box and polyA sites. On E-box DNA, each monomer predominantly formed HBs with one DNA strand, leading to a "biased strand association" (Fig. 5A and Fig. S15, S16A) as identified from monomer TF–DNA association.[47] Early in the simulations (first 10 ns), flanking DNA sequence contacts occurred sporadically (average occupancy ∼50–60%; Fig. S15, top). By 100 ns, HBs in the E-box core stabilized at ∼80% occupancy and remained largely consistent up to 1 μs (Fig. S15, bottom; Fig. 5A).

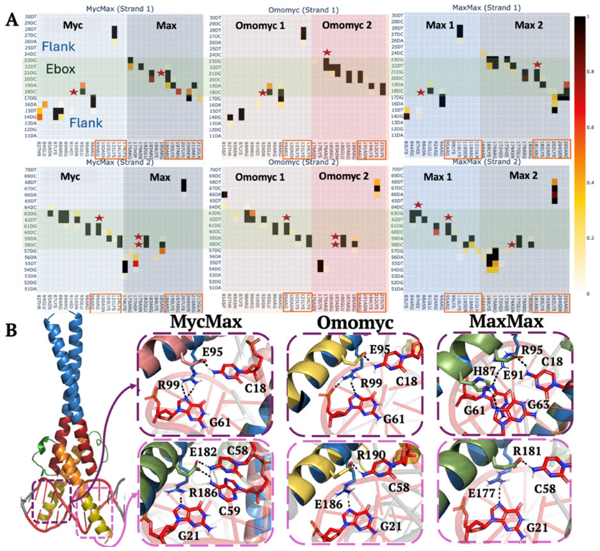

Fig. 5 Protein–DNA hydrogen bond (HB) interactions and patterns in the MycMax, Omomyc dimer, and MaxMax complexes. (A) Heatmaps of HB occupancy for MycMax, Omomyc, and MaxMax complexes interacting with the DNA from individual 1-μs MD simulations. HB contacts with the DNA bases are marked with red stars. The E-box region is highlighted in green, and each protein monomer is delimited by a rectangular colored box. The HLH region is delimited by an orange box on the *x*-axis. (B) Molecular visualization of key protein–DNA HBs. Each monomer is shown in a separate row. The top row shows Myc, Omo1 and Max1, respectively. The bottom row shows Max, Omo2, and Max2, respectively.

Conserved HBs maintain E-box recognition but weaken on polyA. At the residue level, a handful of conserved residues, including arginines and occasionally a lysine, glutamic acid, or asparagine, formed the majority of HBs across all three systems (Fig. S16C and Fig. 5B). Four positions in the BR (E95/R95–C18, R186/E177–G21, E182/R190/R181–C58, and R99/R95–G61) consistently contacted E-box DNA bases in MycMax, the Omomyc dimer, and MaxMax. Even at an atomic level, most contacts with occupancy larger than 15% were conserved across all three systems (Fig. S16). On polyA, however, the overall HB occupancy dropped substantially from 44% to 12% for MycMax, 74% to 58% for Omomyc, and slightly from 20% to 18% for MaxMax, although the DNA strand association bias persisted on polyA (Fig. S16B–S18). Some conserved residues continued to form HBs, but these interactions were less frequent and were no longer specific (*i.e.*, targeted DNA bases) on polyA.

In particular, one can compare HB patterns between Myc (in MycMax) and DNA with those of the Omomyc monomer and DNA, as the BRs presumably in association with DNA are identical in Myc and Omomyc. Interestingly, one can see that although Myc–DNA HB patterns are formed similarly to those of Omomyc1–DNA (Fig. 5A), Omomyc1 forms additional HBs with flank DNA on the preferred strand 2 (on 66DA) while Myc forms more HBs with flank DNA on the non-preferred strand 1 (*e.g.*, on 15DT to 16DA).

These observations highlight a common recognition mechanism centered on a stable E-box core and a system-dependent use of flanking base contacts. MycMax and MaxMax both show more flanking HBs than Omomyc on the E-box (Fig. S17). We also noted a preference for single-strand binding within the E-box core observed as in recently studied monomeric TF domain protein WRKY on W-box,[47,48] which may enhance sequence specificity. This idea was further supported by Kullback–Leibler (KL) divergence analyses (see below).

Statistical measure of TF–DNA HB binding specificity or relative entropy between E-box and polyA. We used Kullback–Leibler[49] (KL) divergence as a statistical measure of relative entropy on protein HB occupancies between different DNA sequences (*e.g.*, E-box and polyA) for stepwise binding specificities, *i.e.*, to quantify how individual HB contacts differ at each DNA nucleotide position for MycMax, the Omomyc dimer, and MaxMax bound to E-box *versus* polyA here. Formally, for a given nucleotide *i*, KL(*i*) is calculated as:

$$\mathrm{KL}(i) = p_{\mathrm{Eb}}^{i} \times \log\left(\frac{p_{\mathrm{Eb}}^{i}}{p_{\mathrm{pol}}^{i}}\right) + \left(1 - p_{\mathrm{Eb}}^{i}\right) \times \log\left(\frac{1 - p_{\mathrm{Eb}}^{i}}{1 - p_{\mathrm{pol}}^{i}}\right) \quad (4)$$

where p_{Eb}^{i} and p_{pol}^{i} denote the probabilities (or occupancies) of forming an HB contact at nucleotide *i* in the E-box and polyA systems, respectively. Larger KL values at a given nucleotide indicate more pronounced differences in the HB contact patterns upon TF binding from one DNA sequence to another, *i.e.*, with higher specificities. Thus, the KL divergence calculated serves as a relative entropy or an information propagation measure to quantify the TF–DNA HB sequence readout

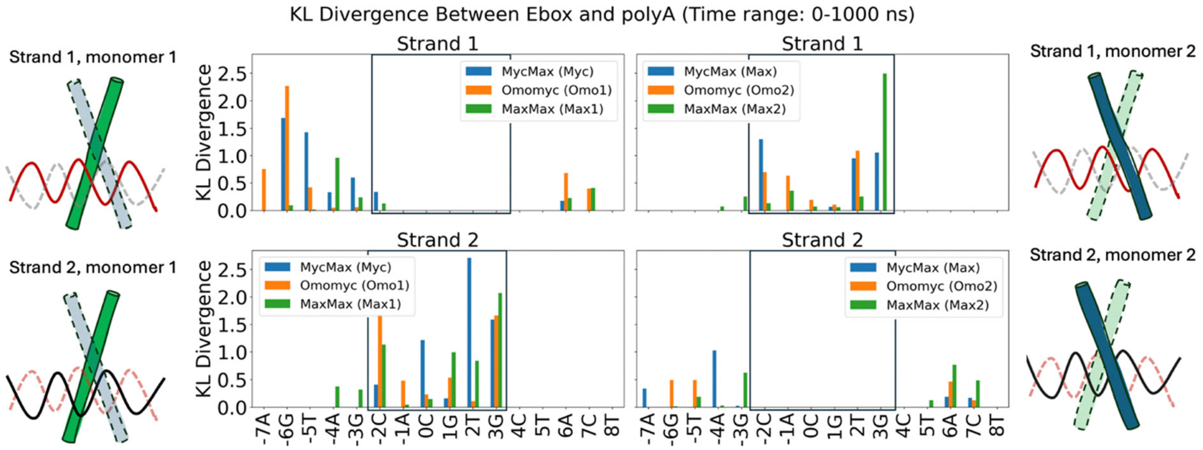

Fig. 6 Comparison of MycMax, Omomyc, and MaxMax HB interactions with E-box and PolyA. KL divergence values, computed over a 1 μs trajectory, compare HB interaction differences between E-box and polyA for MycMax (blue), Omomyc (orange), and MaxMax (green) across both DNA strands. The KL divergence at each nucleotide quantifies how HB occupancies change when the TF binds to polyA *versus* E-box. The E-box region is highlighted by a black box. DNA numbering is re-scaled to have the central C in the E-box CACGTG as index 0. Schematics illustrate the strand preference: monomer 1 preferentially interacts with DNA strand 2 (left), while monomer 2 shows preference for DNA strand 1 (right).

specificities, one nucleotide at a time, here between the highly specific E-box and the homogeneous polyA reference.

Biased strand association enhances DNA sequence readout, with increased sensitivity at E-box edges and flanking sites. In the E-box core, each monomer showed high KL divergences on its preferred DNA strand, reflecting that the advantage of the biased strand association is to enhance the DNA sequence readout (Fig. 6 and 5A). For example, Myc (in MycMax) and Max1 (in MaxMax) consistently had larger KL values at positions 2T and 3G within the CACGTG motif, indicating strong sequence specificity or information readout at these two nucleotide positions. In contrast, the non-preferred strand (*i.e.*, with very few HB contacts) at the same positions showed near-zero KL divergence, suggesting minimal differences in HB contacts or not much DNA sequence information readout there. The central CACGTG nucleotides (positions 0C and 1G) had lower KL values than the right-edge or outer (including the left-edge −2C and −1A) nucleotides, although the base of the central G is a key contributor to high-affinity E-box binding. The prominence of the outer edge nucleotides in sequence readout can be explained by the CANNTG definition of the E-box,[17] where CA and TG steps are recognized as important for specificity, and NN at the center are less constrained for specific recognition.

Outside the 6-bp E-box, certain flanking nucleotides also showed moderate to large KL peaks (KL values ∼1–2), particularly on strand 1, which is predominantly contacted by Myc, Omo1, and Max1 on the left flanking site (Fig. 6, top left).

Overall, the KL divergence analysis confirms that E-box *versus* polyA (Fig. S19–S20) binding sequence specificity or constraint is most pronounced at the edge of the E-box, and the DNA information readout is notable on the DNA strand each monomer prefers in association. Meanwhile, flanking DNA sites also contribute to sequence sensitivity, almost comparable to the specific DNA bound region.

Conclusions

In this study, we used microsecond-scale MD simulations and MMGBSA energy analyses to compare how the three bHLHLZ TFs MycMax, the Omomyc dimer, and MaxMax bind to a canonical E-box (5′-CACGTG-3′) *versus* reference polyA (5′-AAAAAA-3′) DNA. Despite sharing a common bHLHLZ structural architecture, our results show distinct structural dynamic behaviors and binding energetic stabilities among the three TF–DNA complexes.

Our MMGBSA calculations reveal that TF–DNA VDW contacts and the BSA (buried surface area) at the interface dominate the binding free energy in all three dimers. This observation aligns with reports that shape complementarity at the protein–DNA interface can achieve binding specificity.[41,43] Notably, MaxMax emerges as the most stabilized complex energetically on the E-box motif but undergoes the most destabilization on polyA. By contrast, the Omomyc dimer consistently exhibits slightly higher stability energetically than MycMax on both the E-box and polyA sequences.

These trends can be compatible with previous experimental studies on Myc and Max.[9,10,15,16] For example, the Max monomer's high intrinsic affinity and its dimer's high affinity to E-box DNA[15] can explain why MaxMax outperforms MycMax stabilizing on the E-box under certain conditions. Meanwhile, some other experiments also show MycMax outcompeting MaxMax on certain E-box variants.[9] We reason that some discrepancies in experimental findings can be attributed to the construct of Myc (full-length or not) or DNA sequences flanking the E-box that contribute to the binding affinity measurements (Fig. 4C). Various flanking sequences may impact TF dimer stability on DNA differently, allowing MaxMax to outcompete MycMax on some occasions,[15,17,18] or MycMax can emerge as the stronger binder.[9] More importantly, MycMax's dimerization is much stronger than MaxMax's,[9,15] which can translate into higher overall binding affinity or stability of MycMax to E-box DNA than that of MaxMax. To test this, experiments at very high concentrations of TF monomers (to prevent dimer dissociation) can be conducted. Besides, MaxMax can bind less favorably than MycMax to nonspecific DNA, as we show on polyA DNA in the current study. Hence, MaxMax can be overall less stable than MycMax on DNA, particularly with nonspecific binding contexts.

Our simulations also show that the Omomyc dimer retains slightly stronger binding than MycMax on DNA of both the E-box to polyA sequences, consistent with earlier EMSA data.[9] Given that Omomyc and Myc share the same bHLH and BR domains, this enhanced binding appears to be allosteric, conferred from LZ mutations to the bHLH and BR–DNA interface. Interestingly, Omomyc dimer stability on DNA is least affected by the switch from E-box to polyA, and it may exhibit the highest affinity to nonspecific DNA, assuming similarly to polyA in the current study. This robust binding may allow Omomyc to effectively compete with and displace Myc from low-affinity DNA sites. In cancer cells, overexpressed Myc can abnormally occupy weak binding motifs, a phenomenon known as "promoter invasion".[10] By maintaining strong binding even on nonspecific DNA, Omomyc may help prevent this aberrant promoter occupancy. Prior studies[10] similarly demonstrated that Omomyc maintains strong binding to noncanonical ("weak") E-boxes, supporting our observation that Omomyc's robust binding range may underlie its therapeutic efficacy in disrupting Myc-mediated transcription.[9–11]

Our HB analyses confirm that each monomer BR in the dimer predominantly binds to one DNA strand, consistent with the biased DNA strand association reported for the monomeric TF domain protein.[47,48] Notably, the interfacial TF–DNA HB contacts are highly conserved across the three systems, suggesting an evolutionarily preserved recognition mechanism, centered mostly at the arginine residues in each BR. These arginines exhibit multivalent HBs, particularly with guanine bases. Our KL divergence analysis further highlights how each monomer's preferred DNA strand shows the greatest difference on HB patterns between E-box and polyA DNA, *i.e.*, to facilitate DNA sequence readout on one DNA strand, particularly on the E-box edge, though flanking bases also exhibit moderate divergence, underscoring the role of peripheral sequences in refining specificity.

In future studies, systematically varying flanking DNA elements could clarify the molecular basis of TF–DNA sequence sensitivity across diverse genomic contexts. Overall, our findings provide new avenues for rationally modulating bHLHLZ TF–DNA interactions and inform improved strategies to target Myc in cancer, potentially exploiting MaxMax's central role in binding specifically to E-box DNA and Omomyc's ability to outcompete Myc across both canonical and noncanonical sites.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the supplementary information (SI). Supplementary information: PDF of additional plots and measurements (.pdf), PDB files of all the systems studied (.zip), and movies from MD simulations (.zip). See DOI: **https://doi.org/10.1039/d5cp03629a**.

Acknowledgements

CM and JY have been supported by UC Cancer Research Coordinating Committee (CRCC) award #C23CR5636. JY had been supported by the CMCF of UCI *via* the NSF grant DMS1763272 and the Simons Foundation grant #594598, and start-up fund from UCI School of Physical Science. CM has been supported by CMCF fellowship *via* the NSF grant DMS1763272, and the Simons Foundation grant #594598. This work utilized the high-performance computing, data storage, and software tools provided by the Research Cyberinfrastructure Center (RCIC) at the University of California, Irvine (UCI). Anton 2 computer time was provided by the Pittsburgh Supercomputing Center (PSC) on award MCB220008P. The Anton 2 machine at PSC was generously made available by D.E. Shaw Research.

References

1 P. J. Hurlin and J. Huang, The MAX-interacting transcription factor network, *Semin. Cancer Biol.*, 2006, **16**, 265–274.

2 B. Lüscher and J. Vervoorts, Regulation of gene transcription by the oncoprotein MYC, *Gene*, 2012, **494**, 145–160.

3 C. V. Dang, MYC on the path to cancer, *Cell*, 2012, **149**, 22–35.

4 M. Eilers and R. N. Eisenman, Myc's broad reach, *Genes Dev.*, 2008, **22**, 2755–2766.

5 S. K. Madden, A. D. de Araujo, M. Gerhardt, D. P. Fairlie and J. M. Mason, Taking the Myc out of cancer: toward therapeutic strategies to directly inhibit c-Myc, *Mol. Cancer*, 2021, **20**, 3.

6 B. Amati, M. W. Brooks, N. Levy, T. D. Littlewood, G. I. Evan and H. Land, Oncogenic activity of the c-Myc protein requires dimerization with Max, *Cell*, 1993, **72**, 233–245.

7 Y. Pan, P. J. van der Watt and S. A. Kay, E-box binding transcription factors in cancer, *Front. Oncol.*, 2023, **13**, 1223208.

8 L. Soucek, M. Helmer-Citterich, A. Sacco, R. Jucker, G. Cesareni and S. Nasi, Design and properties of a Myc derivative that efficiently homodimerizes, *Oncogene*, 1998, **17**, 2463–2472.

9 L. A. Jung, A. Gebhardt, W. Koelmel, C. P. Ade, S. Walz, J. Kuper, B. Von Eyss, S. Letschert, C. Redel, L. D'Artista, A. Biankin, L. Zender, M. Sauer, E. Wolf, G. Evan, C. Kisker and M. Eilers, OmoMYC blunts promoter invasion by oncogenic MYC to inhibit gene expression characteristic of MYC-dependent tumors, *Oncogene*, 2017, **36**, 1911–1924.

10 M. J. Demma, C. Mapelli, A. Sun, S. Bodea, B. Ruprecht, S. Javaid, D. Wiswell, E. Muise, S. Chen, J. Zelina, F. Orvieto,

A. Santoprete, S. Altezza, F. Tucci, E. Escandon, B. Hall, K. Ray, A. Walji and J. O'Neil, Omomyc Reveals New Mechanisms To Inhibit the MYC Oncogene, *Mol. Cell. Biol.*, 2019, **39**(22), e00248-19.

11 D. Massó-Vallés and L. Soucek, Blocking Myc to Treat Cancer: Reflecting on Two Decades of Omomyc, *Cells*, 2020, **9**(4), 883.

12 E. Garralda, M.-E. Beaulieu, V. Moreno, S. Casacuberta-Serra, S. Martínez-Martín, L. Foradada, G. Alonso, D. Massó-Vallés, S. López-Estévez, T. Jauset, E. Corral de la Fuente, B. Doger, T. Hernández, R. Perez-Lopez, O. Arqués, V. Castillo Cano, J. Morales, J. R. Whitfield, M. Niewel, L. Soucek and E. Calvo, MYC targeting by OMO-103 in solid tumors: a phase 1 trial, *Nat. Med.*, 2024, **30**, 762–771.

13 M. Allevato, E. Bolotin, M. Grossman, D. Mane-Padros, F. M. Sladek and E. Martinez, Sequence-specific DNA binding by MYC/MAX to low-affinity non-E-box motifs, *PLoS One*, 2017, **12**(7), e0180147.

14 P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models, *Acc. Chem. Res.*, 2000, **33**, 889–897.

15 J. Hu, A. Banerjee and D. J. Goss, Assembly of b/HLH/z Proteins c-Myc, Max, and Mad1 with Cognate DNA: Importance of Protein−Protein and Protein−DNA Interactions, *Biochemistry*, 2005, **44**, 11855–11863.

16 O. Ecevit, M. A. Khan and D. J. Goss, Kinetic Analysis of the Interaction of b/HLH/Z Transcription Factors Myc, Max, and Mad with Cognate DNA, *Biochemistry*, 2010, **49**, 2627–2635.

17 M. Allevato, E. Bolotin, M. Grossman, D. Mane-Padros, F. M. Sladek and E. Martinez, Sequence-specific DNA binding by MYC/MAX to low-affinity non-E-box motifs, *PLoS One*, 2017, **12**, e0180147.

18 J. Guo, T. Li, J. Schipper, K. A. Nilson, F. K. Fordjour, J. J. Cooper, R. Gordân and D. H. Price, Sequence specificity incompletely defines the genome-wide occupancy of Myc, *Genome Biol.*, 2014, **15**, 482.

19 S. K. Nair and S. K. Burley, X-ray structures of Myc-Max and Mad-Max recognizing DNA: Molecular bases of regulation by proto-oncogenic transcription factors, *Cell*, 2003, **112**, 193–205.

20 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, The Protein Data Bank, *Nucleic Acids Res.*, 2000, **28**(1), 235–242.

21 A. R. Ferré-D'Amaré, G. C. Prendergast, E. B. Ziff and S. K. Burley, Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain, *Nature*, 1993, **363**, 38–45.

22 UniProt: the Universal Protein Knowledgebase in 2025, Nucleic Acids Research, Oxford Academic, **https://academic.oup.com/nar/article/53/D1/D609/7902999**, (accessed 4 March 2025).

23 L. Schrödinger and W. DeLano, PyMOL, 2020, Available at: **http://www.pymol.org/pymol**.

24 S. Li, W. K. Olson and X.-J. Lu, Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures, *Nucleic Acids Res.*, 2019, **47**, W26–W34.

25 T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations, *Nucleic Acids Res.*, 2004, **32**, W665–W667.

26 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX*, 2015, **1–2**, 19–25.

27 P. Bauer, B. Hess and E. Lindahl, GROMACS 2022 Source Code, 2022, **https://doi.org/10.5281/zenodo.6103835**.

28 D. E. Shaw, J. P. Grossman, J. A. Bank, B. Batson, J. A. Butts, J. C. Chao, M. M. Deneroff, R. O. Dror, A. Even, C. H. Fenton, A. Forte, J. Gagliardo, G. Gill, B. Greskamp, C. R. Ho, D. J. Ierardi, L. Iserovich, J. S. Kuskin, R. H. Larson, T. Layman, L.-S. Lee, A. K. Lerer, C. Li, D. Killebrew, K. M. Mackenzie, S. Y.-H. Mok, M. A. Moraes, R. Mueller, L. J. Nociolo, J. L. Peticolas, T. Quan, D. Ramot, J. K. Salmon, D. P. Scarpazza, U. B. Schafer, N. Siddique, C. W. Snyder, J. Spengler, P. T. P. Tang, M. Theobald, H. Toma, B. Towles, B. Vitale, S. C. Wang and C. Young, *in* SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2014, pp. 41–53.

29 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.

30 Refinement of the Sugar–Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA, *J. Chem. Theory Comput.*, **https://pubs.acs.org/doi/10.1021/acs.jctc.5b00716**, (accessed 3 March 2025).

31 T. Darden, D. York and L. Pedersen, Particle mesh Ewald: An N log(N) method for Ewald sums in large systems, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

32 M. Parrinello and A. Rahman, Crystal Structure and Pair Potentials: A Molecular-Dynamics Study, *Phys. Rev. Lett.*, 1980, **45**, 1196–1199.

33 S. Genheden and U. Ryde, The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities, *Expert Opin. Drug Discovery*, 2015, **10**, 449–461.

34 B. R. I. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, MMPBSA.py: An Efficient Program for End-State Free Energy Calculations, *J. Chem. Theory Comput.*, 2012, **8**, 3314–3321.

35 M. S. Valdés-Tresanco, M. E. Valdés-Tresanco, P. A. Valiente and E. Moreno, Gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS, *J. Chem. Theory Comput.*, 2021, **17**, 6281–6291.

36 Interaction Entropy: A New Paradigm for Highly Efficient and Reliable Computation of Protein–Ligand Binding Free Energy, *J. Am. Chem. Soc.*, **https://pubs.acs.org/doi/10.1021/jacs.6b02682**, (accessed 20 February 2025).

37 L. Duan, G. Feng, X. Wang, L. Wang and Q. Zhang, Effect of electrostatic polarization and bridging water on

CDK2–ligand binding affinities calculated using a highly efficient interaction entropy method, *Phys. Chem. Chem. Phys.*, 2017, **19**, 10140–10152.

38 G. D. Hawkins, C. J. Cramer and D. G. Truhlar, Pairwise solute descreening of solute charges from a dielectric medium, *Chem. Phys. Lett.*, 1995, **246**, 122–129.

39 V. Tsui and D. A. Case, Theory and applications of the generalized born solvation model in macromolecular simulations, *Biopolymers*, 2000, **56**, 275–291.

40 T. Tuccinardi, What is the current value of MM/PBSA and MM/GBSA methods in drug discovery?, *Expert Opin. Drug Discovery*, 2021, **16**, 1233–1237.

41 R. Rohs, X. Jin, S. M. West, R. Joshi, B. Honig and R. S. Mann, Origins of specificity in protein–DNA recognition, *Annu. Rev. Biochem.*, 2010, **79**, 233–269.

42 B. Jayaram, K. McConnell, S. B. Dixit, A. Das and D. L. Beveridge, Free-energy component analysis of 40 protein–DNA complexes: A consensus view on the thermodynamics of binding at the molecular level, *J. Comput. Chem.*, 2002, **23**, 1–14.

43 R. Rohs, S. M. West, A. Sosinsky, P. Liu, R. S. Mann and B. Honig, The role of DNA shape in protein–DNA recognition, *Nature*, 2009, **461**, 1248–1253.

44 R. Gordân, N. Shen, I. Dror, T. Zhou, J. Horton, R. Rohs and M. L. Bulyk, Genomic Regions Flanking E-Box Binding Sites Influence DNA Binding Specificity of bHLH Transcription Factors through DNA Shape, *Cell Rep.*, 2013, **3**, 1093–1104.

45 N. M. Luscombe, R. A. Laskowski and J. M. Thornton, Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level, *Nucleic Acids Res.*, 2001, **29**, 2860–2874.

46 J. DeRouchey, B. Hoover and D. C. Rau, A Comparison of DNA Compaction by Arginine and Lysine Peptides: A Physical Basis for Arginine Rich Protamines, *Biochemistry*, 2013, **52**, 3000–3009.

47 Revealing atomic-scale molecular diffusion of a plant-transcription factor WRKY domain protein along DNA, *PNAS*, **https://www.pnas.org/doi/abs/10.1073/pnas.2102621118?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub++0pubmed**, (accessed 20 February 2025).

48 Nonspecific *vs.* specific DNA binding free energetics of a transcription factor domain protein, *Biophys. J.*, **https://www.cell.com/biophysj/fulltext/S0006-3495(23)00661-6**, (accessed 20 February 2025).

49 S. Kullback and R. A. Leibler, On Information and Sufficiency, *Ann. Math. Stat.*, 1951, **22**, 79–86.