Mechanism of cognate sequence discrimination by the ETS-family transcription factor ETS-1

Received for publication, February 1, 2019, and in revised form, May 1, 2019. Published, Papers in Press, May 2, 2019. DOI 10.1074/jbc.RA119.007866

Kenneth Huang‡1, Suela Xhani‡1, Amanda V. Albrecht‡, Van L. T. Ha§, Shingo Esaki‡, and © Gregory M. K. Poon‡§2

From the ⌧Department of Chemistry and ⌧Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, Georgia 30303

Edited by Norma M. Allewell

Functional evidence increasingly implicates low-affinity DNA recognition by transcription factors as a general mechanism for the spatiotemporal control of developmental genes. Although the DNA sequence requirements for affinity are well-defined, the dynamic mechanisms that execute cognate recognition are much less resolved. To address this gap, here we examined ETS-1, a paradigm developmental transcription factor, as a model for which cognate discrimination remains enigmatic. Using molecular dynamics simulations, we interrogated the DNA-binding domain of murine ETS1 alone and when bound to high- and low-affinity cognate sites or to nonspecific DNA. The results of our analyses revealed collective backbone and side-chain motions that distinguished cognate versus nonspecific as well as high- versus low-affinity cognate DNA binding. Combined with binding experiments with site-directed ETS1 mutants, the molecular dynamics data disclosed a triad of residues that respond specifically to low-affinity cognate DNA. We found that a DNA-contacting residue (Gln-336) specifically recognizes low-affinity DNA and triggers the loss of a distal salt bridge (Glu-343/Arg-378) via a large side-chain motion that compromises the hydrophobic packing of two core helices. As an intact Glu-343/Arg-378 bridge is the default state in unbound ETS1 and maintained in high-affinity and nonspecific complexes, the low-affinity complex represents a unique conformational adaptation to the suboptimization of developmental enhancers.

Developing organisms depend on the expression of essential genes in a spatially and temporally coordinated manner. Functional studies in model animals, including sea squirts (1, 2), Drosophila spp. (3–5), Caenorhabditis elegans (6), sea stars and sea urchins (7), and the annelid Platynereis dumerilii (8), show that low-affinity binding by transcription factors is essential for developmental gene regulation. Low-affinity (“suboptimal”) binding enables spatiotemporal control of genes by tuning the sensitivity of their enhancers to transcription factor levels and facilitating combinatorial control by multiple factors (9, 10). Substituting the WT low-affinity sites with high-affinity sequences disrupts the level, location, timing, and specificity of developmentally sensitive genes with major dysmorphic outcomes. As the literature increasingly implicates low-affinity transcription factor–DNA complexes as key mediators of gene expression, mechanisms of sequence discrimination are re-emerging as deserving areas of detailed investigation.

To this end, the ETS family of transcription factors represents an excellent model system for tackling this line of inquiry. ETS proteins arose early in metazoan evolution (11, 12) and are ubiquitous in animals (13, 14). Despite significant divergence in both function and primary sequence, ETS proteins harbor an eponymous DNA-binding domain that is highly conserved. The ETS domain is a winged helix–loop–helix motif that recognizes a helical turn of DNA sequences harboring a core 5′-GG(A/T)-3′ consensus. Beyond this specific requirement, ETS domains recognize a wide palette of flanking-base variants (and their epigenetic modifications) with a dispersion of affinities (15–17). ETS complexes formed with consensus-bearing (cognate) sequences are experimentally distinguishable from nonspecific complexes, quantitatively in terms of affinity and qualitatively by distinctive DNase I hypersensitivity in ETS-bound DNA at the core consensus (18, 19). Cognate discrimination is a biologically relevant function because the genomic distribution of ETS-bound DNA sequences in vivo corresponds closely to the relative affinities for their respective ETS domains (20). ETS members that are more stringent site discriminators in binding experiments are, in turn, more selective in their genomic motifs (21, 22). Finally, low-affinity ETS binding was specifically required for proper activation of enhancers governing the patterning of the anterior neural plate in Ciona embryos (1).

Structurally, ETS domains are characterized by a characteristic segmented pattern of contacts with cognate DNA. Although contacts with the core motif are mediated with nucleobases, flanking contacts occur exclusively by interactions with backbone phosphate or deoxyribose oxygen atoms. Low-affinity binding therefore appears to arise from variation in the core-flanking contacts. Since proposing this “indirect readout” mechanism of cognate discrimination (16), there has been limited progress in elucidating its molecular basis. In particular, little is known about how ETS domains dynamically parse cogn...
nate targets. NMR studies have established that the ETS domains of ETS1 (23) and ETV6 (24) utilize the same interface in both cognate and nonspecific DNA complexes. Nonetheless, the precise interactions that distinguish “low-affinity” and “nonspecific” complexes remain obscure. This is a critical distinction because nonspecific DNA dominates cognate sites in abundance but does not constitute bona fide binding sites for transcription factor function.

To address these questions, we selected ETS1, a paradigm ETS-family regulator of key developmental processes such as hematopoiesis (25) and cardiac neural crest development (26, 27) as a model system. Heteronuclear NMR studies of ETS1 have shown that high-affinity and nonspecific DNA binding invoke large differences in 15N spin-lattice and spin-spin relaxation in the bond vectors as well as amide 15N NOE of the bound protein (23). As these parameters reflect conformational dynamics on the picosecond to nanosecond timescale, we hypothesize that molecular dynamics (MD) simulations should provide useful high-resolution information regarding the target-dependent differences in binding affinity. The current state of highly optimized biomolecular force fields, particularly for DNA (28), further support unbiased MD as a viable approach. We therefore compared explicit solvent simulations of the ETS domain of murine ETS1 (residues 331–440) in the unbound state and in complex with high-affinity, low-affinity cognate (consensus-bearing), or nonspecific DNA. These simulations represent the longest-duration and broadest-scope all-atom studies of any ETS protein to date (29–33). The results are consistent with literature data and reveal novel mechanisms of sequence discrimination. Additional simulations and experiments with mutants prompted by observations on WT ETS1 identified specific electrostatic contacts distal from the protein–DNA interface as important determinants of high-versus low-affinity cognate binding. The results have implications for the evolution of low-affinity binding in developmental enhancers as well as autoinhibition, a general mechanism of allosteric regulation and protein–protein partnerships by transcription factors (34).

Results

To maximize the relevance of the simulations to experiments, a binary co-crystal structure (PDB code 1K79) that most closely corresponded to the C-terminal ETS domain (ΔN331) used in experimental studies (15, 17, 21, 35–37) was used to template the simulational models. For the complexes, the crystallographic DNA sequence was replaced with DNA duplexes utilized in experimental determinations of the binding affinity under the same conditions (0.15 M NaCl at 25 °C, Fig. 1). The high- and low-affinity cognate sequences (termed in the literature SC1 and SC12) were established substrates originally identified in SELEX (systematic evolution of ligands by exponential enrichment) screening against this same ETS1 construct (15). Binding experiments indicated that the high- and low-affinity cognate complexes of ΔN331 differed in stability by 13-fold (ΔG* = +6.40 kJ/mol) and that the nonspecific complex bound to a randomized sequence was an additional 28-fold weaker (ΔG* = +8.27 kJ/mol) than the low-affinity complex.

The established secondary structure assignment for ETS domains, a winged helix–loop–helix motif, is H1–S1–S2–H2–loop–H3–S3–wing–S4 (H, α-helix; S, β-strand). H3 is the essential recognition helix that inserts into the major groove of consensus DNA and is the most conserved element in the ETS domain. As in native ETS1, ΔN331 contained two short C-terminal helices, H4 and H5, that are not part of the canonical ETS topology (38). However, ΔN331 did not include additional N-terminal helices (termed HI-1 and HI-2) that are extrinsic to the winged helix–loop–helix motif and negatively regulate DNA binding. Each structure was run in triplicate independent simulations seeded with random initial velocities. Equilibration was judged by the root mean square deviation of the protein and that the nonspecific complex bound to a randomized sequence was an additional 28-fold weaker (ΔG* = +8.27 kJ/mol) than the low-affinity complex.

The established secondary structure assignment for ETS domains, a winged helix–loop–helix motif, is H1–S1–S2–H2–loop–H3–S3–wing–S4 (H, α-helix; S, β-strand). H3 is the essential recognition helix that inserts into the major groove of consensus DNA and is the most conserved element in the ETS domain. As in native ETS1, ΔN331 contained two short C-terminal helices, H4 and H5, that are not part of the canonical ETS topology (38). However, ΔN331 did not include additional N-terminal helices (termed HI-1 and HI-2) that are extrinsic to the winged helix–loop–helix motif and negatively regulate DNA binding. Each structure was run in triplicate independent simulations seeded with random initial velocities. Equilibration was judged by the root mean square deviation of the protein and that the nonspecific complex bound to a randomized sequence was an additional 28-fold weaker (ΔG* = +8.27 kJ/mol) than the low-affinity complex.

Validation of simulations against experiments

Our first objective was to establish whether the simulations generated reasonable models of ETS1 compared with experimental data. Relative to the co-crystallographic template, the bound proteins relaxed without major changes in secondary structure (HA, 1.9%; LA, 2.8%; NS, 0.9%; Fig. S2A). The
unbound state showed larger differences (8.6%) relative to the closest solution NMR apo structure (ΔN301, PDB code 1R36) that harbors the additional N-terminal helices HI-1 and HI-2 (39). Differences were localized at H1, the unnamed single-turn helix between H1 and S1, S3, and H5, which were all close to the HI helices. Notably, ΔN331 from the co-crystal structure also showed a 9% difference in backbone/side-chain RMSF content relative to the NMR apo structure (Fig. S2B). In the absence of an experimental structure of unbound ΔN331, the secondary structure difference between the unbound models appeared to reflect conformational contributions from the additional residues in ΔN301.

\(^{15}N\) NOE experiments showed that the backbone amides of ETS1 near and at the DNA interface became less dynamic on the picosecond to nanosecond timescale upon binding DNA (23). Comparison of root mean square fluctuations (RMSF) between the bound and unbound states showed reduced backbone dynamics in all simulational bound structures as well (Fig. 2A). The decreased dynamics mapped principally to the loop (between H2 and the recognition helix H3), H3, and the wing (between S3 and S4) (Fig. 2B). The nonspecific complex retained significantly greater fluctuations relative to specific complexes, also in agreement with the NOE data (23). With respect to the side chains, the difference in RMSF of nonhydrogen side-chain atoms between unbound and bound ETS1 indicated that most DNA-contacting residues were also stabilized when bound (Fig. 2C). Also in agreement with a reported comparison of bound and free NMR structures (37), several residues displayed increased dynamics relative to unbound ETS1. Thus, the general agreement between experiments and simulations by multiple structural and dynamic metrics supports the simulations as viable models.

As an additional check on the models, we tested whether the unbound and bound states were reversible on the simulation timescale (10\(^{-7}\) s). To do so, we exchanged the unbound and
bound protein states at the final frame of their respective original simulations. New unbound systems were generated by deleting the DNA from bound systems, and new bound systems were constructed by mapping the first converged unbound protein onto DNA. Positional restraints were applied to the DNA backbone for the first 700 ns to preserve the bound conformation. Re-equilibration of the newly unbound and bound systems, as judged by root mean square deviation, occurred from 1.3 to 1.5 μs, and the final 300 ns was used for analysis. We compared the equilibrated trajectories from the original and exchange simulations by principal component analysis (PCA) of all nonhydrogen atoms from the combined trajectories (Fig. S3). In each case, corresponding pairs of original and exchanged trajectories overlapped along the first two principal components, indicating highly similar conformations. Thus, even though the simulations were unbiased and did not sample longer-timescale dynamics, the distinct dynamics of the various states were reversible on the nanosecond to microsecond timescale, supporting the physical relevance of the simulations in interpreting experimental binding data.

Comparison of the sequence-dependent protein–DNA interfaces formed by ETS1

As a parsimonious initial approach to comparing the interactions in the ETS1 complexes, we enumerated hydrogen bonds (direct and water-mediated) and charge–charge interactions between the protein and DNA along standard criteria as described under “Experimental procedures.” Overall, nonspecifically bound ETS1 made fewer DNA contacts than any cognate sequence, particularly in direct hydrogen bonds and charge–charge interactions (Table S1). We then scrutinized three residues (Leu-337, Arg-391, and Arg-394) that have been experimentally implicated as critical for cognate binding (23, 40). Leu-337 at the N terminus of H1 contacts DNA via a backbone–backbone contact that is nevertheless sensitive to side-chain mutations (40). Arg-391 and Arg-394 are signature residues in the recognition helix H3, which are absolutely conserved in all ETS domains and responsible for interacting with the core 5′-GGAA(T)/A-3′ consensus.

In the B-type interfaces, residues corresponding to the loop and wing plus several residues in helix H4 (residues 378–383) and the wing (residues 404–408) adjoining S3 and S4 (Fig. 3B). Both the loop and wing interact with bases flanking the core consensus. Projected along PC1, this motion corresponded to the loop and wing moving toward (negative scores) and away from the DNA (Fig. 3C). The distributions of motion along PC1 qualitatively differentiated (in sign) cognate and nonspecific interactions. In contrast with PC1, the collective motion corresponding to PC2 was diffusely distributed between the loop and wing plus several residues in helix H4 (Fig. S6, A and B). Projection along PC2, which captured collective motions that were orthogonal to PC1, did not meaningfully differentiate the bound and unbound states (Fig. 3C). This residual collective backbone motion therefore described intrinsic dynamics and stochastic noise that were independent of DNA.

Because the dominant collective backbone motion in PC1 corresponded to the loop and wing moving in and out of proximity to DNA, we examined the density of protein–DNA contacts along these elements in the three complexes. Progression from positive to negative along PC1 (i.e. nonspecific > high-affinity > low-affinity) corresponded to an increase in hydro-
gen bonds between protein and DNA along the loop, particularly direct protein-to-DNA contacts (Fig. S7). This trend followed the global enumeration of contacts (Table S1) which showed a higher density of contacts in the low-affinity complex over the high-affinity one. Both approaches therefore indicated that backbone motion did not account for high- and low-affinity discrimination, further motivating examination of side-chain dynamics.

The scree plot for the side-chain PCA showed that the collective side-chain dynamics were more directed than for the backbone, wherein the first PC alone accounted for >50% of the total variance (Fig. 4A). All but one of the dominant residues contributing to side-chain PC1 were located in the protein–DNA interface, spanning in sequence from Arg-378 in helix H2 to the end of the wing (Fig. 4B). The lone residue outside the DNA contact surface is Glu-343 in H1, which is proximal to Arg-378 in H2. Globally, the motion defined by side-chain PC1 described a concerted contraction or expansion of the protein in correspondence to an ordered change in solvent-accessible surface area (SASA) of the bound protein by ~10% (Fig. S8A). The distributions of fluctuations showed that collective side-chain motion along PC1 completely distinguished the high- and low-affinity complexes (Fig. 4C) whereas backbone PC1 did not (cf. Fig. 3C). Nonspecific complexes remained overlapped with the unbound state.

Although dominant loop and wing residues in side-chain PC1 were also among those in backbone PC1, their quantitative contributions did not correspond significantly, reflective of uncorrelated components in backbone and side-chain dynamics hinted at by the RMSF data (cf. Fig. 2). In addition, several other residues contributed significantly to side-chain PC1 but not to backbone PC1. Of particular interest were Arg-378, the most dominant contributor to PC1, and Glu-343, located in H1 and well-separated from the DNA interface. Their side chains showed a concerted range of motion along PC1 from being in close proximity to far apart (Fig. S9). Residual fluctuations along side-chain PC2, which were distributed over a similar range of residues (Fig. S6, C and D), did not segregate the four states (Fig. 4C) and showed no trend in SASA of the bound protein (Fig. S8B).

Summarizing the principal component analyses, collective motions of all four states for ETS1 were primarily captured by the respective first PC for the backbone and side chains. Projection of the backbone dynamics along this PC separated cognate from nonspecific binding. Cognate sites, but not nonspecific DNA, induced backbone motions at the structural elements flanking the recognition helix toward the DNA. This accounted for the higher density of DNA contacts in cognate complexes over their nonspecific counterparts. Projection of side-chain dynamics along its first PC separated high- and low-affinity cognate binding. Collective side-chain motion was dominated by a charged residue, Arg-378 in helix H2, that underwent a large coupled motion with a complementary partner, Glu-343 in H1. As the concerted dynamics of Arg-378 and Glu-343 presented themselves as prominent contrasting features in the high- versus low-affinity complex, we examined their contributions to DNA binding in greater detail.

Cognate sequence selection is coupled allosterically to a distal salt bridge

Because side-chain PC1 completely distinguished high- and low-affinity binding, we scrutinized the conformations of Arg-
378, the most dominant contributor to PC1, and its partner Glu-343 in the two cognate complexes (Fig. 5A). In the high-affinity complex, the side chains of Arg-378 and Glu-343 formed a persistent salt bridge with a time-average center-of-mass separation of $4.0 \pm 0.1$ Å (Fig. 5B), noncovalently stapling H1 and H2 together. In the low-affinity complex, Arg-378 instead contacted a DNA backbone phosphate downstream of the 5′-GGAA-3′ consensus. This was a dramatic change in local conformation, as the side chains were separated by $15.1 \pm 0.1$ Å in the low-affinity complex.

In addition to the E343–R378 interaction, the low-affinity complex was characterized by a large reduction in DNA backbone contact by Gln-336, a residue proximal to H1 (Fig. 5A). Gln-336 and Arg-378 had been identified as two residues with significant differences in DNA contacts between the high- and low-affinity complexes (Fig. S5). In the high-affinity complex, the side-chain amide nitrogen of Gln-336 maintained a time-averaged distance of $3.7 \pm 0.3$ Å to the closest phosphate oxygen. In the low-affinity complex, the Gln-336 side chain was mostly solvent-exposed at $6.2 \pm 0.1$ Å from the DNA. To interrogate the triad of Gln-336, Glu-343, and Arg-378 more definitively, we carried out a side-chain PCA consisting of these three residues only. The resulting correlation uniquely separated the low-affinity complex from all other states (Fig. S10). The low-affinity complex was uniquely characterized by concerted side-chain dynamics in which contact by Gln-336 with the DNA backbone adjacent to the core consensus was allosterically coupled to the distal Glu-343/Arg-378 salt bridge. Because Gln-336 resided at the DNA contact surface, the results suggested that Gln-336 was directly sensitive to DNA identity, directing the fate of the Glu-343/Arg-378 salt bridge whose disruption specifically flagged a low-affinity complex.

To further clarify these relationships, we simulated a panel of complexes harboring various point mutations at Glu-343 or Gln-336. Of several mutations tested at one or the other position, only high-affinity complexes of E343L and Q336L perturbed the dynamics and significantly sampled PC space occupied by the WT low-affinity complex (Fig. S11). The Glu-343/Arg-378 distance profiles of WT and mutant complexes showed that the salt bridge was predominantly intact in both high- and low-affinity mutant Q336L complexes, unlike their WT counterparts (Fig. 5B). Thus, discrimination between high- and low-affinity sequences, insofar as the Glu-343/Arg-378 salt bridge was concerned, was abolished in the Q336L mutant. For E343L, elimination of the salt bridge rendered high-affinity binding similar to low-affinity binding by WT ETS1. Low-affinity binding by E343L was even more perturbed, with no well-defined Q336L–DNA contact (Fig. 5C). DNA contact by Arg-378 in both E343L complexes were also sharply reduced. In summary, the simulations of the mutants corroborated the WT results. These simulations reinforced the Glu-343/Arg-378 salt bridge as an important contributor to high-affinity binding in WT ETS1 and Gln-336 as a relay of sequence information to the salt bridge.

To test this hypothesis experimentally, we cloned E343L and Q336L mutants of ΔN331 and measured their affinities for high- and low-affinity DNA (Fig. 5D). E343L bound both high- and low-affinity DNA more poorly than WT ETS1 but retained target selectivity (Fig. 5E). The reduction in affinity in both mutants was consistent with the Glu-343/Arg-378 salt bridge as an important contributor to high-affinity binding in WT ETS1 and Gln-336 as a relay of sequence information to the salt bridge.
mised, reflective of Gln-336 as a sensor of high- versus low-affinity DNA. Q336L bound both cognate sequences similarly within experimental precision (Fig. 5D). Notably, Q336L showed significantly higher affinities relative to the low-affinity E343L complex. The binding experiments thus supported Gln-336 as a selectivity sensor that triggered the loss of a distal stabilizing salt bridge (Glu-343/Arg-378) upon binding to low-affinity DNA. This sensor operated only in the low-affinity complex, as an intact Glu-343/Arg-378 salt bridge was the default state in unbound ETS1 and maintained in both high-affinity and nonspecific complexes.

To interrogate the triadic residues in a more functional context, we examined the Mnx enhancer, which regulates development of the notochord in Ciona embryos (1, 2). A key 5′ regulatory region in the Mnx enhancer harbored three cognate ETS sites, one of which is high-affinity and the other two weak (Fig. 6A). The low-affinity (suboptimal) site proximal to the 5′-terminus is absolutely required for correctly patterned Mnx expression in vivo (2). Alteration of this site to a high-affinity sequence (“optimized ETS,” or OE, in keeping with the reported nomenclature) results in strong but ectopic Mnx expression outside of the notochord (2). A mechanistic link between cognate discrimination and functional embryogenesis would therefore predict differences in Mnx occupancy by WT ETS1 and the selection-competent mutant E343L but not the incompetent mutant Q336L. We titrated native (“Mnx/H1102bp”) and OE enhancers with WT and mutant ETS1 and resolved the various complexes (unbound and up to triply bound states) by native gel electrophoresis (Fig. 6B). Across identical ranges in protein concentrations, both E343L and Q336L bound the enhancers more weakly than WT (Fig. 6B, arrows). Crucially, at protein concentrations that just saturated the enhancer sites, both WT and E343L were qualitatively different in their relative occupancies of the available cognate sites. WT and E343L exhibited differential preferences for the singly and doubly bound states. In contrast, Q336L showed the same relative occupancy and therefore failed to discriminate the two enhancers. The data therefore strongly supported cognate discrimination by ETS1 and its basis in the triadic residues as functionally important in chordate embryogenesis.

Breach of the Glu-343/Arg-378 salt bridge compromises the structural integrity of H2 and exposes the hydrophobic core of H1

In their salt-bridged configuration, the sidechains of Glu-343 and Arg-378 spanned over significant portions of the H1 and H2 helices. The hydrophobic composition of H1 (LWQFLLELLT, Glu-343 underlined) suggested that loss of the Glu-343/Arg-378 salt bridge, which is intact in unbound ETS1, would be detrimental to protein stability. To test this hypothesis, we analyzed the conformations of H1 and H2 in the simulations models. First, differences in solvent exposure were
determined in terms of accessible water at each residue in H1 (residues 337–345) and H2 (368–378) relative to the unbound protein (Fig. 7A). In the high-affinity complex, in which the salt bridge was intact, Leu-337 and Trp-375 were less water-exposed, on average, relative to unbound ETS1. Loss of the salt bridge in the low-affinity complex resulted in exposure of several hydrophobic residues to water in H1 (Phe-340, Glu-343) and H2 (Gly-376), whereas Arg-378 became less water-exposed upon burial near the DNA (cf. Fig. 6A). H2 was well-preserved in both high- and low-affinity complexes up to Gly-376 (Fig. 7B), beyond which backbone H-bond contacts became less favorable in a DNA-dependent manner. As a result, helicity in the low-affinity complex was significantly compromised (Fig. 7C). As Arg-378 resided at the C terminus of H2, these results suggest that a salt bridge with Glu-343 significantly stabilized the local tertiary structure. The presence of a strong helix breaker like glycine in H2 may serve as a leverage point to facilitate the contortion of Arg-378 in H2.

**Discussion**

**A conserved triad of residues mediates low-affinity binding by ETS1**

The side-chain dynamics exhibited by the triad of Gln-336/Glu-343/Arg-378 in WT complexes qualitatively distinguished low-affinity binding from interactions with optimal and nonspecific DNA. These results were experimentally verified by the testing of post hoc hypotheses on the simulational mutations E343L and Q336L. To our knowledge, this triad represents the first specific elucidation of how interactions with backbone contacts at core-flanking positions are translated to differential affinity in a eukaryotic transcription factor. This study therefore extends our interpretation of the cache of literature data on DNA site selection with novel insights into an area of transcription factor biophysics that has seen limited progress in recent years.

More broadly, identification of this triad contributes to our understanding of how suboptimization of developmental enhancers arises evolutionarily. Although the cis-regulatory components of suboptimized enhancers are increasingly defined (9, 10, 41), the molecular adaptations in trans-regulatory proteins remain unknown. An intact Glu-348/Arg-378 salt bridge is the default state in unbound ETS1 and staunchly maintained in both high-affinity and nonspecific binding. This observation logically argues that low-affinity binding, which uniquely disrupts this salt bridge to incur an affinity penalty, represents an evolved molecular trait. If “de-tuning” of DNA sites constituted an evolutionary mechanism for achieving spatiotemporal control of developmental genes, the present data argue for a parallel adaptation in their trans-regulatory partners to suboptimized DNA. In ETS1, frustration introduced by low-affinity DNA to destabilize packing of the H2 helix represents such a mechanism. These distortions are facilitated by a glycine (Gly-376) in H2, a highly disfavored residue in α-helices. The conservation of Gly-376 at corresponding positions in H2

---

**Figure 6. Cognate discrimination by ETS1 at the Mnx enhancer, a cis-regulatory element in embryogenesis.** A, sequence and schematic of the Mnx enhancer in a syntax that produces correctly patterned expression in Ciona embryos (Mnx −2bp), or a variation (Mnx −2bp OE) that causes ectopic expression in vivo. The presentation maintains consistency with the reported nomenclature (2). Blue symbols correspond to ETS-binding sites, with the shading and number denoting the affinity relative to an optimal site (such as SC1). The red symbol represents a binding site for ZicL, another required transcription factor, that overlaps with the central weak ETS site. B, representative titrations of each enhancer with WT, selection-competent E343L, and the incompetent Q336L mutant, as resolved gel mobility shift. Each protein-containing lane represents a 1.5-fold change in concentration. Arrows indicate the 5% bound threshold in each titration. To the right of each gel is the lane trace for the rightmost lane, representing the distribution of bound states at a just-saturating concentration of protein. The numbers represent unbound (0) and bound states harboring 1 to 3 equivalents of protein.
Suboptimization of ETS1 binding

Figure 7. Breach of a key salt bridge compromises the hydrophobic core of ETS1. A, time-averaged water contacts within a 3.4 Å radius for each residue in WT ETS1 near H1 and H2 (orange, 333–378). The change in solvent exposure was calculated as the difference from unbound protein. Values below an assigned cutoff (≤ 1 water) are colored gray and those above black. Residues with values above/below the cutoff in H1 and H2 are marked with an asterisk. High-affinity binding resulted in loss of water exposure at Leu-337 (H1) and Trp-375 (H2). In contrast, the low-affinity binding increased water exposure at Phe-340, Glu-343 (H1), and Gly-376 (H2). Arg-378 became less hydrated because of DNA contact. B, local structure of helix H2, showing the backbone conformation from Arg-374 and Arg-378. C, values indicating the fraction of frames in which H2 residues were helical as scored by DSSP.

across ETS family proteins therefore hints at a possible shared component of cognate discrimination (42). Moreover, the triad of Glu-343/Arg-378/Gln-336 that mediates cognate discrimination is strongly conserved in ETS1 orthologs across vast evolutionary distances, from simple water-borne organisms to higher mammals (Fig. S12). Finally, harking to a role of ETS proteins as developmental gene regulators, this triad is present in several putative proto-ETS proteins in Amphimedon queenslandica, a model species of sponge, one of the earliest phyla of animals with a defined body plan (43).

Role of sequence discrimination on self-regulation and protein–protein partnerships

In addition to DNA binding, the ETS domain is known to mediate regulatory interactions with other modules in ETS1 (23, 44, 45). Two extrinsic helices (HI-1 and HI-2) flanking the N terminus of the ETS domain interact with its C-terminal helices H4 and H5 (46). These interactions generate an autoinhibited state (Fig. 8A) in which unfolding of the HI helices is coupled to DNA binding by H1, specifically Gln-336 and Leu-337 at its N terminus (31, 40). The present data extend this model by accounting for the DNA dependence of autoinhibition, which has heretofore been unexplained. Paired data on binding by the uninhibited and autoinhibited ETS domain to variants to the high-affinity cognate SC1 (17, 40, 48) demonstrate that binding to weaker cognates is more strongly autoinhibited than higher-affinity sequences in a graded manner. Quantitatively, the free energy relationship between binding by autoinhibited versus minimal ETS1 yields a slope significantly above unity (Fig. 8B). The proportionate autoinhibition in binding methylated SC1 is consistent with established functional inhibition DNA methylation on ETS1 transactivation in vivo (49). Remarkably, the correlation applies whether the perturbation at a consensus-flanking position involves a change in base sequence, CpG methylation, or even a single-stranded nick and seems not to depend on the perturbed position itself. The literature data therefore indicate that autoinhibition depends on intrinsic affinity, which, in the context of the model, must be communicated from the DNA contact surface to the autoinhibitory helices via H1.
The absence of the Glu-343/Arg-378 salt bridge in low-affinity complexes would imply the loss of a hydrogen bond from Ala-323 to the side chain of Glu-343 on H1, whereas the latter is salt-bridged with Arg-378 (50). We propose that, in low-affinity binding, loss of the Glu-343/Arg-378 salt bridge (triggered by the DNA sensor Gln-336) also destabilizes the Glu-343/Ala-323 contact and packing of HI-2, causing it to further unfold in addition to HI-1 (Fig. 8C). More complete unraveling of both HIs would, in turn, increase the activation penalty to low-affinity DNA binding by autoinhibited ETS1. The triadic residues may therefore present a mechanism by which the severity of autoinhibition can be tuned in response to the nature of the DNA substrate.

This extended model of ETS1 autoinhibition poses important implications in terms of protein–protein partnerships known to exert combinatorial control of Ets1-dependent genes. Binding partners such as Pax5 (51, 52) and Runx1 (48, 53) or homodimerization with another ETS1 (50, 54) activate ETS1 activity by relieving autoinhibition. In the ternary Pax5–ETS1–DNA co-crystal structure (PDB code 1MDM), for example, HI-2 is in a folded state as the protein is bound to an intrinsically nonconsensus site for ETS1 (55). We may apply our extended model and interpret the Pax5 partnership in terms of mitigating the penalty of unfolding HI-2. This interpretation is further enhanced by a dedicated side chain–side chain interaction between Pax5 (Gln-22) with Gln-336 of ETS1 in the ternary complex, which may bypass the trigger that would otherwise disrupt the Glu-343/Arg-378 salt bridge.

Additional long timescale (microsecond to millisecond) motions have been inferred from model-free analysis of heteronuclear spin relaxation data and directly by relaxation dispersion in unbound ETS1 (45, 46, 56). These slow dynamics link the ETS domain to the autoinhibitory modules and indicate at least two active/inactive conformations in equilibrium without DNA. The labile active form can sample affinity-dependent interactions, which are less accessible to the rigid inactive state. Oscillations between these states may also aid in diffusion on nonspecific sequences (56). However, this behavior is inverted upon binding DNA. High-affinity sites stabilize short and long timescale motions at the DNA interface and HIs, respectively, more relative to lower-affinity sequences (23). Similar to the ETS domain, the dynamic properties of longer ETS1 constructs likely diverge upon binding DNA. It is therefore likely that fast dynamics, such as the Glu-343/Arg-378 salt bridge, modulate the slower motion that ultimately determines the DNA-dependent variation in inhibitory penalties.

In conclusion, we have determined the molecular mechanism of cognate discrimination by ETS1, a model developmental gene regulator. Conventionally, sequence-specific DNA recognition is focused on interactions at the protein–DNA interface, where concepts such as DNA shape have garnered strong traction in the literature (57–59). Site selectivity is regarded as a continuum from nonspecific DNA on one hand to optimal sequences on the other, along which low-affinity sequences represent imperfect intermediates. The present combination of computational and experimental data suggests a modified perspective, however, as the low-affinity complex of ETS1 exhibits dynamics, mediated by a Gln-336/Glu-343/Arg-378 triad, that are qualitatively distinct from the nonspecifically and high-affinity bound state with distinct impact on allosteric autoinhibition. Given the ancient history of ETS factors and their structural homology to the myriad variations of the helix–turn–helix motif, the parsimonious interpretation is that a

**Figure 8. Proposed mechanism of variable autoinhibition in ETS1.** A, DNA binding by ETS1 is autoinhibited by an allosteric mechanism mediated by helix H1 (blue), which couples the DNA binding interface with an N-terminal inhibitory module consisting of the short helices HI-1 and HI-2 (red). Structural evidence shows that both HI helices are folded in the unbound state (PDB code 1R36, left). Upon binding to high-affinity cognate DNA (right; PDB code 3MFK, the DNA sequence is 5′-GCCGAGAAGTG-3′, consensus in orange), and HI-1 is differentially unfolded over HI-2, free energy relationship between ETS1 binding to the same DNA substrates as uninhibited (i.e. ΔN331) and autoinhibited (i.e. ΔN280) constructs. Data are compiled from the literature (see text for references) in decreasing uninhibited affinity as follows: 1, SC1; 2 and 3, nicked SC1 (green); 4, 6–8, and 10, flanking base variants of SC1; 5, a long construct harboring SC1; 9 and 11, hemi- and fully methylated SC1 (magenta). Reported error bars are shown. The line represents an error-weighted linear fit by York’s method with the 95% confidence interval. The estimated slope is 1.5 ± 0.1. Note the base-10 logarithmic axes. The log (K_C, uninhibited) / log (K_C, autoinhibited) line is unity. The present results suggest that the absence of the Glu-343/Arg-378 salt bridge in low-affinity complexes would destabilize HI-2, leading to increased autoinhibition (arrows).

How may this affinity inhibition scaling come about? Previous work has shown that the dynamics of Gln-336 and Leu-337 act together as a “conformational switch” to direct correlated motions between H1 and H4 in autoinhibited ETS1 (31). The simulations show that Leu-337 interacts with DNA comparably in both high- and low-affinity complexes (HA, 67%; LA, 72%). Thus, Leu-337 is an important residue in this scheme, but it is not completely disordered, nor are they equally unstructured DNA show that the autoinhibitory helices HI-1 and HI-2 are more structured than the distal HI-1, as evidenced by chemical shift perturbations (23), spin relaxation dynamics (23), and amide proton exchange rates (46). In comparable co-crystallographic high-affinity cognate complexes (50), HI-2 is helical, whereas HI-1 adopts a coil conformation. HI-2 maintains a backbone hydrogen bond from Ala-323 to the side chain of Glu-343 on H1, whereas the latter is salt-bridged with Arg-378 (50). We propose that, in low-affinity binding, loss of the Glu-343/Arg-378 salt bridge (triggered by the DNA sensor Gln-336) also destabilizes the Glu-343/Ala-323 contact and packing of HI-2, causing it to further unfold in addition to HI-1 (Fig. 8C). More complete unraveling of both HIs would, in turn, increase the activation penalty to low-affinity DNA binding by autoinhibited ETS1. The triadic residues may therefore present a mechanism by which the severity of autoinhibition can be tuned in response to the nature of the DNA substrate.

This extended model of ETS1 autoinhibition poses important implications in terms of protein–protein partnerships known to exert combinatorial control of Ets1-dependent genes. Binding partners such as Pax5 (51, 52) and Runx1 (48, 53) or homodimerization with another ETS1 (50, 54) activate ETS1 activity by relieving autoinhibition. In the ternary Pax5–ETS1–DNA co-crystal structure (PDB code 1MDM), for example, HI-2 is in a folded state as the protein is bound to an intrinsically nonconsensus site for ETS1 (55). We may apply our extended model and interpret the Pax5 partnership in terms of mitigating the penalty of unfolding HI-2. This interpretation is further enhanced by a dedicated side chain–side chain interaction between Pax5 (Gln-22) with Gln-336 of ETS1 in the ternary complex, which may bypass the trigger that would otherwise disrupt the Glu-343/Arg-378 salt bridge.

Additional long timescale (microsecond to millisecond) motions have been inferred from model-free analysis of heteronuclear spin relaxation data and directly by relaxation dispersion in unbound ETS1 (45, 46, 56). These slow dynamics link the ETS domain to the autoinhibitory modules and indicate at least two active/inactive conformations in equilibrium without DNA. The labile active form can sample affinity-dependent interactions, which are less accessible to the rigid inactive state. Oscillations between these states may also aid in diffusion on nonspecific sequences (56). However, this behavior is inverted upon binding DNA. High-affinity sites stabilize short and long timescale motions at the DNA interface and HIs, respectively, more relative to lower-affinity sequences (23).

Similar to the ETS domain, the dynamic properties of longer ETS1 constructs likely diverge upon binding DNA. It is therefore likely that fast dynamics, such as the Glu-343/Arg-378 salt bridge, modulate the slower motion that ultimately determines the DNA-dependent variation in inhibitory penalties.

In conclusion, we have determined the molecular mechanism of cognate discrimination by ETS1, a model developmental gene regulator. Conventionally, sequence-specific DNA recognition is focused on interactions at the protein–DNA interface, where concepts such as DNA shape have garnered strong traction in the literature (57–59). Site selectivity is regarded as a continuum from nonspecific DNA on one hand to optimal sequences on the other, along which low-affinity sequences represent imperfect intermediates. The present combination of computational and experimental data suggests a modified perspective, however, as the low-affinity complex of ETS1 exhibits dynamics, mediated by a Gln-336/Glu-343/Arg-378 triad, that are qualitatively distinct from the nonspecifically and high-affinity bound state with distinct impact on allosteric autoinhibition. Given the ancient history of ETS factors and their structural homology to the myriad variations of the helix–turn–helix motif, the parsimonious interpretation is that a
**Suboptimization of ETS1 binding**

Noncontinuum mechanism of cognate discrimination is not unique to ETS1. Pervading evidence of functionally essential low-affinity sites in the enhancer elements of developmental genes provides independent support for this view and a renewed need for improved understanding of how eukaryotic transcription factors parse cognate variants. As ETS1 shows, molecular dynamics simulations of cognate complexes offer an attractive approach to addressing the mechanics of enhancer site suboptimization, particularly in the absence of appropriate experimental structures.

**Experimental procedures**

**Molecular dynamics simulation**

Initial coordinates were constructed by modeling the WT or mutant ETS domain of murine ETS1 (residues 331 to 440, termed ΔN331) to the protein in a cognate ETS1–DNA co-crystal structure (PDB code 1K79) (55). The protonation states of all ionizable residues were estimated using PROPKA3 (60) and set accordingly for pH 7.4. Unbound (apo) ETS1 was generated by deleting the DNA from the co-crystal structure. DNA-bound complexes were produced by replacing the crystallographic sequence with a 10-bp DNA sequence inserted into a 5′-CGGCAA…ATGGCG-3′ cassette to generate a 23-bp construct. Unbound DNA models were generated as B-form DNA by NAB. All systems were constructed in tLeap by placing the solute in a truncated octahedral TIP3P waterbox with a minimum distance of 10 Å to the box edge. Excess ions (Na+, Cl−) were added to achieve a salt concentration of 0.15 M. Topologies were generated using the Amber14SB and OL15 force fields for protein and DNA, respectively (28, 61). All energy minimization and molecular dynamics simulations were performed with AMBER16. The particle mesh Ewald method was used for treating long-range electrostatic interactions with a 10 Å cutoff for nonbonded interactions. All bonds involving hydrogen atoms were constrained by the SHAKE algorithm. All systems were energy-minimized using steepest descent (2,000 steps) followed by conjugate gradient (1,000 steps). Positional restraints on all nonsolvent atoms were used (500 kcal/mol Å2) before being released stepwise for unrestrained minimization. The systems were then heated to 300 K over 100 ps in the canonical ensemble using the Langevin thermostat with restraints of different solute groups in the following order: protein side chain, followed by 1 ns of equilibration without restraints of all ionizable residues were estimated using PROPKA3 (60) and set accordingly for pH 7.4. Unbound (apo) ETS1 was generated by deleting the DNA from the co-crystal structure. DNA-bound complexes were produced by replacing the crystallographic sequence with a 10-bp DNA sequence inserted into a 5′-CGGCAA…ATGGCG-3′ cassette to generate a 23-bp construct. Unbound DNA models were generated as B-form DNA by NAB. All systems were constructed in tLeap by placing the solute in a truncated octahedral TIP3P waterbox with a minimum distance of 10 Å to the box edge. Excess ions (Na+, Cl−) were added to achieve a salt concentration of 0.15 M. Topologies were generated using the Amber14SB and OL15 force fields for protein and DNA, respectively (28, 61). All energy minimization and molecular dynamics simulations were performed with AMBER16. The particle mesh Ewald method was used for treating long-range electrostatic interactions with a 10 Å cutoff for nonbonded interactions. All bonds involving hydrogen atoms were constrained by the SHAKE algorithm. All systems were energy-minimized using steepest descent (2,000 steps) followed by conjugate gradient (1,000 steps). Positional restraints on all nonsolvent atoms were used (500 kcal/mol Å2) before being released stepwise for unrestrained minimization. The systems were then heated to 300 K over 100 ps in the canonical ensemble using the Langevin thermostat with restraints on the solute (100 kcal/mol Å2). Afterward, the isobaric–isothermal ensemble was thermalized at 1 atm for 8 ns total in 1-ns steps. Initial relaxation of solute restraints (75, 50, 25, and 10 kcal/mol Å2) were followed by an iterative release of restraints of different solute groups in the following order: nucleobases, nucleotide backbone, protein backbone, and protein side chain, followed by 1 ns of equilibration without restraints. A time step of 1 fs was used for both heating and equilibration. Subsequent production runs were run for up to 1.5 μs with a 2-fs time step, saving frames every 4 ps.

**MD trajectory analysis**

Analysis of the trajectories was performed using ccpptraj from AmberTools16. Visualization was performed with PyMOL. The final 300 ns (75,000 frames) was used for analysis. The energy-minimized structure was used as a common reference for least square fitting. Root mean square deviations were calculated for all heavy atoms in the protein and complex. Root mean square fluctuations were calculated on a per-residue basis separately for the backbone atoms (Cα, C, N, and O) and non-hydrogen side-chain atoms of the protein. Hydrogen bond analysis was performed using a 3.0 Å distance cutoff and an angle cutoff of 135° for direct and bridging (one water) contacts. Charge–charge interactions were enumerated for charge-bearing atoms of charged residues of the protein (Lys, NZ; Arg, NH1 and NH2) within a predefined cutoff (3.5, 5.5, and 7.5 Å) from the phosphate oxygen atoms (OP1 and OP2) of the DNA. SASA was calculated for all atoms of the solute with a probe radius of 1.4 Å. Collective motions in the protein in bound and unbound states were determined by PCA of backbone or nonhydrogen side-chain atoms for the systems as a single concatenated trajectory. Because of the high flexibility of the N and C termini, the first and last three residues (333–335 and 436–438) were omitted in PCA. Block averages were calculated by splitting the entire dataset into five equally sized blocks and computing the mean for each block before determining the mean and standard deviation across the blocks.

**Molecular cloning and recombinant protein preparation**

Site-specific mutants of the murine ETS1 ETS domain (residues 331 to 440, termed ΔN331) were generated by PCR cloning as described previously (37). Briefly, His6-tagged recombinant constructs of WT and mutant ΔN331 were overexpressed in *Escherichia coli* and purified by Co-NTA affinity chromatography as described previously (37) to apparent homoogeneity as judged by SDS-PAGE. Following cleavage of the affinity tag with thrombin, purified protein was extensively dialyzed against 10 mM Tris-HCl (pH 7.4) containing 0.15 M NaCl and 3 mM DTT.

**ETS1–DNA binding experiments**

The free energy of DNA binding by WT and mutant ETS1 was determined via the equilibrium constant by fluorescence polarization, as reported previously (22). In brief, a subsaturating concentration of WT or mutant ΔN331 in complex with a Cy3-labeled DNA probe was displaced at 25 °C with unlabeled DNA duplexes harboring the same cognate sequences as the simulational DNA. The solution conditions were 10 mM Tris-HCl (pH 7.4) containing 0.1% BSA, 0.15 M NaCl, and 3 mM freshly prepared DTT. Steady-state fluorescence anisotropy of the probe was measured at 530/590 nm with a Molecular Devices Paradigm plate reader. The binding data were fitted by nonlinear least square analysis with a competitive binding model (17), as a function of total unlabeled DNA concentration, to estimate the equilibrium dissociation constant *K*<sub>D</sub>. The free energy of binding was computed via the standard thermodynamic relationship *ΔG*<sup>‡</sup> = *RT* ln *K*<sub>D</sub>. Titration samples involving Mnx enhancers were prepared in a direct binding format using DNA PCR-amplified from plasmids harboring the fragment sequence. These plasmids were prepared by cloning synthetic oligos encoding the NCBI reference sequence NW_004190369.2 into pCR2.1-TOPO vectors (Invitrogen). Amplicons were purified by agarose electrophoresis and extracted with spin columns (Thermo). Each sample consisted of 10 nM and graded concentrations of purified WT
or mutant (E343L or Q336L) in the same buffer matrix as samples for fluorescence polarization. The mixtures were resolved by native electrophoresis in 10% polyacrylamide gel running at 10 V/cm in 0.5× Tris/glycine buffer. Gels were stained with SYBR Gold (Thermo) and imaged with UV transillumination. Each lane was quantified by peak fitting with Gaussian distributions.

Author contributions—K. H., S. X., and G. M. K. P. conceptualization; K. H. data curation; K. H. software; K. H., S. X., A. V. A., V. L. T. H., S. E., and G. M. K. P. formal analysis; K. H., S. X., and G. M. K. P. validation; K. H., S. X., A. V. A., V. L. T. H., S. E., and G. M. K. P. investigation; K. H. visualization; K. H., S. X., V. L. T. H., and G. M. K. P. methodology; K. H., S. X., and G. M. K. P. writing—original draft; K. H., S. X., A. V. A., V. L. T. H., S. E., and G. M. K. P. writing—review and editing; S. X. and G. M. K. P. supervision; G. M. K. P. resources; G. M. K. P. funding acquisition; G. M. K. P. project administration.

References

1. Farley, E. K., Olson, K. M., Zhang, W., Brandt, A. J., Rokhsar, D. S., and Levine, M. S. (2015) Suboptimization of developmental enhancers. Science 350, 325–328 CrossRef Medline
2. Farley, E. K., Olson, K. M., Zhang, W., Rokhsar, D. S., and Levine, M. S. (2016) Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. Proc. Natl. Acad. Sci. U.S.A. 113, 6508–6513 CrossRef Medline
3. Jiang, J., and Levine, M. (1993) Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. Cell 72, 741–752 CrossRef Medline
4. Crocker, J., Abe, N., Rinaldi, L., McGregor, A. P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F., Mann, R. S., and Stern, D. L. (2015) Low affinity binding site clusters confer hox specificity and regulatory robustness. Cell 160, 191–203 CrossRef Medline
5. Ramos, A. I., and Barolo, S. (2013) Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. Philos. Trans. R Soc. Lond. B. Biol. Sci. 368, 20130018 CrossRef Medline
6. Gaudet, J., and Mango, S. E. (2002) Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. Science 295, 821–825 CrossRef Medline
7. Cary, G. A., Cheatte Jarvela, A. M., Francolini, R. D., and Hinman, V. F. (2017) Genome-wide use of high- and low-affinity Tbrain transcription factor binding sites during ecdinolmedi development. Proc. Natl. Acad. Sci. U.S.A. 114, 5854–5861 CrossRef Medline
8. Handberg-Thorsager, M., Gutierrez-Mazariego, J., Arold, S. T., Kumar Nadendla, E., Bertucci, P. Y., Germain, P., Tomancek, P., Pierzchalski, K., Jones, J. W., Albatal, R., Kane, M. A., Bourguet, W., Lautet, V., Arendt, D., and Schubert, M. (2018) The ancestral retinoic acid receptor was a low-affinity sensor triggering neuronal differentiation. Sci. Adv. 4, eaao1261 CrossRef Medline
9. Crocker, J., Noon, E. P., and Stern, D. L. (2016) The soft touch: low-affinity transcription factor binding sites in development and evolution.Curr. Top. Dev. Biol. 117, 455–469 CrossRef Medline
10. Weingarten-Gabbay, S., and Segal, E. (2014) The grammar of transcriptional regulation. Hum. Genet. 133, 701–711 CrossRef Medline
11. Degnan, B. M., Degnan, S. M., Nagarajna, T., and Morse, D. E. (1993) The ETS multigene family is conserved throughout the metazoa. Nucleic Acids Res. 21, 3479–3484 CrossRef Medline
12. Lautot, V., Niel, C., Dutereque-Coquillaud, M., Leprince, D., and Stehelin, D. (1993) Evolution of the ETS gene family. Biochem. Biophys. Res. Commun. 190, 8–14 CrossRef Medline
13. Wang, Z., and Zhang, Q. (2009) Genome-wide identification and evolutionary analysis of the animal specific ETS transcription factor family. Evol. Bioinform. Online 5, 119–131 CrossRef Medline
14. Lautot, V., Hänni, C., Stéhelin, D., and Dutereque-Coquillaud, M. (1999) Molecular phylogeny of the ETS gene family. Oncogene 18, 1351–1359 CrossRef Medline
15. Nye, J. A., Petersen, J. M., Gunther, C. V., Jonsen, M. D., and Graves, B. J. (1992) Interaction of murine ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. Genes Dev. 6, 975–990 CrossRef Medline
16. Szmyczyna, B. R., and Arrowsmith, C. H. (2000) DNA binding specificity studies of four ETS proteins support an indirect read-out mechanism of protein-DNA recognition. J. Biol. Chem. 275, 28363–28370 CrossRef Medline
17. Stephens, D. C., and Poon, G. M. (2016) Differential sensitivity to methylated DNA by ETS-family transcription factors is intrinsically encoded in their DNA-binding domains. Nucleic Acids Res. 44, 8671–8681 CrossRef Medline
18. Graves, B. J., Gillespie, M. E., and McIntosh, L. P. (1996) DNA binding by the ETS domain. Nature 384, 322 CrossRef Medline
19. Poon, G. M. (2012) Sequence discrimination by DNA-binding domain of ETS family transcription factor PU.1 is linked to specific hydration of protein-DNA interface. J. Biol. Chem. 287, 18297–18307 CrossRef Medline
20. Wei, G. H., Badis, G., Berger, M. F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A. R., Yan, J., Talukder, S., Turunen, M., Taipale, M., Stunnenberg, H. G., et al. (2010) Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. EMBO J. 29, 2147–2160 CrossRef Medline
21. He, G., Tolic, A., Bashkin, J. K., and Poon, G. M. (2015) Heterogeneous dynamics in DNA site discrimination by the structurally homologous DNA-binding domains of ETS-family transcription factors. Nucleic Acids Res. 43, 4322–4331 CrossRef Medline
22. Albrecht, A. V., Kim, H. M., and Poon, G. M. K. (2018) Mapping interfacial hydration in ETS-family transcription factor complexes with DNA: a chimeric approach. Nucleic Acids Res. 46, 10577–10588 CrossRef Medline
23. Desjardins, G., Okon, M., Graves, B. J., and McIntosh, L. P. (2016) Conformational dynamics and the binding of specific and nonspecific DNA by the autoinhibited transcription factor Ets-1. Biochemistry 55, 4105–4118 CrossRef Medline
24. De, S., Chan, A. C., Coyne, H. J., 3rd, Bhachec, N., Hermsdorf, U., Okon, M., Murphy, M. E., Graves, B. J., and McIntosh, L. P. (2014) Steric mechanism of auto-inhibitory regulation of specific and non-specific DNA binding by the ETS transcriptional repressor ETV6. J. Mol. Biol. 426, 1390–1406 CrossRef Medline
25. Ciau-Ulitz, A., Wang, L., Patient, R., and Liu, F. (2013) ETS transcription factors in hematopoietic stem cell development. Blood Cells Mol. Dis. 51, 248–255 CrossRef Medline
26. Gao, Z., Kim, G. H., Mackinnon, A. C., Flagg, A. E., Bassett, B., Earley, J. U., and Svensson, E. C. (2010) Ets1 is required for proper migration and differentiation of the cardiac neural crest. Development 137, 1543–1551 CrossRef Medline
27. Wang, C., Kam, R. K., Shi, W., Xia, Y., Chen, X., Cao, Y., Sun, J., Du, Y., Lu, G., Chen, Z., Chan, W. Y., Chan, S. O., Deng, Y., and Zhao, H. (2015) The proto-oncogene transcription factor EtsI regulates neural crest development through histone deacetylase 1 to mediate output of bone morphogenetic protein signaling. J. Biol. Chem. 290, 21925–21938 CrossRef Medline
28. Galindo-Murillo, R., Robertson, J. C., Zagarová, M., Šponer, J., Otyepka, M., Jurečka, P., and Cheatham, T. E., 3rd (2016) Assessing the current state of Amber force field modifications for DNA. J. Chem. Theory Comput. 12, 4114–4127 CrossRef Medline
29. Obika, S., Reddy, S. Y., and Bruce, T. C. (2003) Sequence-specific DNA binding of Ets-1 transcription factor: molecular dynamics study on the Ets domain-DNA complexes. J. Mol. Biol. 331, 345–359 CrossRef Medline
30. Reddy, S. Y., Obika, S., and Bruce, T. C. (2003) Conformations and dynamics of Ets-1 ETS-domain-DNA complexes. Proc. Natl. Acad. Sci. U.S.A. 100, 15475–15480 CrossRef Medline
31. Kamberaj, H., and van der Vaart, A. (2009) Correlated motions and interactions at the onset of the DNA-induced partial unfolding of Ets-1. Biophys. J. 96, 1307–1317 CrossRef Medline
Suboptimization of ETS1 binding

32. Kamberaj, H., and van der Vaart, A. (2009) Extracting the causality of correlated motions from molecular dynamics simulations. *Biophys. J.* **97**, 1747–1755 CrossRef Medline

33. Karolak, A., and van der Vaart, A. (2012) Importance of local interactions for the stability of inhibitory helix 1 in apo Ets-1. *Biophys. Chem.* **165–166**, 74–78 CrossRef Medline

34. Pufall, M. A., and Graves, B. J. (2002) Autoinhibitory domains: modular effectors of cellular regulation. *Annu. Rev. Cell Dev. Biol.* **18**, 421–462 CrossRef Medline

35. Jonsen, M. D., Petersen, J. M., Xu, Q. P., and Graves, B. J. (1996) Characterization of the cooperative function of inhibitory sequences in Ets-1. *Mol. Cell Biol.* **16**, 2065–2073 CrossRef Medline

36. Wang, S., Linde, M. H., Munde, M., Carvalho, W. D., Wilson, W. D., and Poon, G. M. (2014) Mechanistic heterogeneity in site recognition by the structurally homologous DNA-binding domains of the ETS family transcription factors Ets-1 and PU.1. *J. Biol. Chem.* **289**, 21605–21616 CrossRef Medline

37. Xhani, S., Esaki, S., Huang, K., Erlitzki, N., and Poon, G. M. (2017) Distinct roles for interfacial hydration in site-specific DNA recognition by ETS-family transcription factors. *J. Phys. Chem. B* **121**, 2748–2758 CrossRef Medline

38. Hollenhorst, P. C., McIntosh, L. P., and Graves, B. J. (2011) Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu. Rev. Biochem.* **80**, 437–471 CrossRef Medline

39. Donaldson, L. W., Petersen, J. M., Graves, B. J., and McIntosh, L. P. (1996) Solution structure of the ETS domain from murren Ets-1: a winged helix-turn-helix DNA binding motif. *EMBO J.* **15**, 125–134 CrossRef Medline

40. Wang, H., McIntosh, L. P., and Graves, B. J. (2002) Inhibitory module of Ets-1 allosterically regulates DNA binding through a dipole-facilitated phosphate contact. *J. Biol. Chem.* **277**, 2225–2233 CrossRef Medline

41. Curina, A., Termanini, A., Barozzi, I., Prosperini, E., Simonatto, M., Polletti, S., Silvola, A., Soldi, M., Austenaa, L., Bonaldi, T., Ghisletti, S., and Natoli, G. (2017) High constitutive activity of a broad panel of housekeeping family transcription factors. *J. Biol. Chem.* **277**, 437–447 CrossRef Medline

42. Zhu, X., Stéhelin, D., Aumercier, M., and van der Vaart, A. (2009) Importance of local interactions and tissue-specific effectors of cellular regulation. *J. Biol. Chem.* **284**, 1248–1253 CrossRef Medline

43. Pufall, M. A., Lee, G. M., Nelson, M. L., Kang, H. S., Velyvis, A., Kay, L. E., and McIntosh, L. P. (2005) The structural and dynamic basis of Ets-1 DNA binding autoinhibition. *J. Biol. Chem.* **280**, 7088–7099 CrossRef Medline

44. Petersen, J. M., Skalický, J. J., Donaldson, L. W., McIntosh, L. P., Alber, T., and Graves, B. J. (1995) Modulation of transcription factor Ets-1 DNA binding: DNA-induced unfolding of an α helix. *Science* **269**, 1866–1869 CrossRef Medline

45. Goetz, T. L., Gu, T. L., Speck, N. A., and Graves, B. J. (2000) Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor α2. *Mol. Cell Biol.* **20**, 81–90 CrossRef Medline

46. Hogart, A., Lichtenberg, I., Ajay, S. S., Anderson, S., NIH Intramural Sequencing Center, Margules, E. H., and Bodine, D. M. (2012) Genome-wide DNA methylation profiles in hematopoietic stem and progenitor cells reveal overrepresentation of ETS transcription factor binding sites. *Genome Res.* **22**, 1407–1418 CrossRef Medline

47. Babayeva, N. D., Wilder, P. J., Shima, M., Mino, K., Desler, M., Ogata, K., Rizzino, A., and Tahirov, T. H. (2010) Structural basis of Ets1 cooperative binding to palindromic sequences on stromelysin-1 promoter DNA. *Cell Cycle* **9**, 3054–3062 CrossRefMedline

48. Garvie, C. W., Pufall, M. A., Graves, B. J., and Wolberger, C. (2002) Structural analysis of the autoinhibition of Ets-1 and its role in protein partnerships. *J. Biol. Chem.* **277**, 45529–45536 CrossRef Medline

49. Fitzsimmons, D., Lukin, K., Lutz, R., Garvie, C. W., Wolberger, C., and Haman, J. (2009) Highly cooperative recruitment of Ets-1 and release of autoinhibition by Pax5. *J. Mol. Biol.* **392**, 452–464 CrossRef Medline

50. Shrivastava, T., Mino, K., Babayeva, N. D., Baranovskaya, O. I., Rizzino, A., and Tahirov, T. H. (2014) Structural basis of Ets1 activation by Runx1. *Leukemia* **28**, 2040–2048 CrossRef Medline

51. Baillat, D., Bégue, A., Stéhelin, D., and Aumercier, M. (2002) ETS-1 transcription factor binds cooperatively to the palindromic head to head ETS-binding sites of the stromelysin-1 promoter by counteracting autoinhibition. *J. Biol. Chem.* **277**, 29386–29398 CrossRef Medline

52. Garvie, C. W., Haman, J., and Wolberger, C. (2001) Structural studies of Ets-1/Pax5 complex formation on DNA. *Mol. Cell* **8**, 1267–1276 CrossRef Medline

53. Pufall, M. A., Lee, G. M., Nelson, M. L., Kang, H. S., Velyvis, A., Kay, L. E., McIntosh, L. P., and Graves, B. J. (2005) Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. *Science* **309**, 142–145 CrossRef Medline

54. Zentner, G. E., Kasinathan, S., Xin, B., Rohs, R., and Henikoff, S. (2015) ChEC-seq kinetics discriminates transcription factor binding sites by DNA sequence and shape in vivo. *Nat. Commun.* **6**, 8733 CrossRef Medline

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

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

57. Olson, M. H., Sondergaard, C. R., Rostkowski, M., and Jensen, J. H. (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7**, 525–537 CrossRef Medline

58. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., et al. (2010) The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature* **466**, 720–726 CrossRef Medline

59. Desjardins, G., Meeker, C. A., Bhachec, N., Currie, S. L., Okon, M., Graves, B. J., and McIntosh, L. P. (2014) Synergy of aromatic residues and phosphoserines within the intrinsically disordered DNA-binding inhibitory elements of the Ets-1 transcription factor. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11019–11024 CrossRef Medline

60. Lee, G. M., Donaldson, L. W., Pufall, M. A., Kang, H. S., Pot, I., Graves, B. J., and McIntosh, L. P. (2008) The affinity of Ets-1 for DNA is modulated by phosphorylation through transient interactions of an unstructured region. *J. Mol. Biol.* **382**, 1014–1030 CrossRef Medline

61. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 CrossRef Medline