Published online 6 February 2017

Nucleic Acids Research, 2017, Vol. 45, No. 5 ***2223–2241***
doi: 10.1093/nar/gkx068

NAR Breakthrough Article

Structured and disordered regions cooperatively mediate DNA-binding autoinhibition of ETS factors ETV1, ETV4 and ETV5

Simon L. Currie[1,2,†], Desmond K. W. Lau[3,†], Jedediah J. Doane[1,2], Frank G. Whitby[4], Mark Okon[3], Lawrence P. McIntosh[3] and Barbara J. Graves[1,2,5,*]

[1]Department of Oncological Sciences, University of Utah, Salt Lake City, UT 84112-5550, USA, [2]Huntsman Cancer Institute, University of Utah, Salt Lake City, UT 84112-5550, USA, [3]Department of Biochemistry and Molecular Biology, Department of Chemistry, and Michael Smith Laboratories, University of British Columbia, Vancouver BC, V6T 1Z3, Canada, [4]Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, UT 84112-5650, USA and [5]Howard Hughes Medical Institute, Chevy Chase, MD 20815-6789, USA

Received November 01, 2016; Revised January 17, 2017; Editorial Decision January 18, 2017; Accepted January 29, 2017

ABSTRACT

Autoinhibition enables spatial and temporal regulation of cellular processes by coupling protein activity to surrounding conditions, often via protein partnerships or signaling pathways. We report the molecular basis of DNA-binding autoinhibition of ETS transcription factors ETV1, ETV4 and ETV5, which are often overexpressed in prostate cancer. Inhibitory elements that cooperate to repress DNA binding were identified in regions N- and C-terminal of the ETS domain. Crystal structures of these three factors revealed an α-helix in the C-terminal inhibitory domain that packs against the ETS domain and perturbs the conformation of its DNA-recognition helix. Nuclear magnetic resonance spectroscopy demonstrated that the N-terminal inhibitory domain (NID) is intrinsically disordered, yet utilizes transient intramolecular interactions with the DNA-recognition helix of the ETS domain to mediate autoinhibition. Acetylation of selected lysines within the NID activates DNA binding. This investigation revealed a distinctive mechanism for DNA-binding autoinhibition in the ETV1/4/5 subfamily involving a network of intramolecular interactions not present in other ETS factors. These distinguishing inhibitory elements provide a platform through which cellular triggers, such as protein–protein interactions or post-translational modifications, may specifically regulate the function of these oncogenic proteins.

INTRODUCTION

Autoinhibition occurs in diverse proteins and allows for spatiotemporal modulation of biological processes in response to various inputs such as signaling pathways and macromolecular interactions (1). This self-dampening behavior can influence the equilibria between the active and inactive states of proteins by serving as the integration point for post-translational modifications and protein partnerships. Partaking in alternative intramolecular and intermolecular interactions is often the key attribute for an autoinhibitory element (2,3). Both structured regions with dynamic character and intrinsically disordered regions (IDRs) can be effective inhibitory elements (4). Conformational flexibility, and even disorder, allows for distinct and adaptable recognition of intramolecular interfaces, as well as surfaces on diverse interacting proteins (5).

The *ETS* gene family, which encodes 28 human transcription factors, has provided a model system to develop an understanding of autoinhibition of sequence-specific DNA binding (6). The conserved ETS domain is autoinhibited in several family members, yet by different mechanisms. For example, a serine-rich IDR allosterically inhibits DNA binding of ETS1 through transient phosphorylation-enhanced interactions with the structured ETS domain and

*To whom correspondence should be addressed. Tel: +1 301 215 8718; Fax: +1 301 215 8828; Email: Barbara.Graves@hci.utah.edu
†These authors contributed equally to the paper as first authors.

Disclaimer: The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIH or other funding agencies.

flanking N- and C-terminal inhibitory α-helices (7,8). In contrast, a single flanking C-terminal α-helix sterically inhibits DNA binding of ETV6 (9–11). In the biological context, autoinhibition of a particular ETS factor provides distinct routes to specific regulation, such as via post-translational modifications (8,12) and protein-protein interactions (13–16). Collectively, this has led to the hypothesis that divergent modes of autoinhibition involving regions flanking the ETS domain help enable specific gene regulation by individual ETS factors (6).

The involvement of the *ETS* genes of the ERG and ETV1/4/5 (also known as PEA3) subfamilies in prostate cancer motivated our goal to expand the mechanistic understanding of autoinhibition to these ETS factors. Chromosomal rearrangements involving the ERG and ETV1/4/5 subfamilies are observed in the majority of prostate cancer tumors (17–19). There is aberrant expression of these full-length, or nearly full-length, ETS proteins upon rearrangement with a prostate-specific or a constitutively expressed promoter (20). In addition, ETV1 and ETV4 mediate PI3-kinase and Ras signaling pathways, resulting in aggressive and metastatic disease phenotypes (21,22). Previous studies suggested that the ETV1/4/5 subfamily also displays autoinhibition of DNA binding (13,23,24); however, detailed characterization, including structural mapping of inhibitory elements and mechanistic insights are lacking. We propose that a full understanding of the autoinhibition of ERG and ETV1/4/5 and its regulation by cellular processes will enable insights into the roles of these factors in prostate cancer progression and provide windows of opportunity for targeted therapeutic interventions.

In this study we describe the molecular basis of DNA-binding autoinhibition in the ETV1/4/5 subfamily of ETS factors. Using ETV4 as a model for this subfamily, we found that inhibitory domains reside both N- and C-terminal of the ETS domain and cooperate to inhibit DNA binding. Crystal structures identified the C-terminal inhibitory domain (CID) as an α-helix that packs against the ETS domain and perturbs the relative positioning of its DNA-recognition helix. Nuclear magnetic resonance (NMR) spectroscopy demonstrated that the N-terminal inhibitory domain (NID) is an IDR that transiently interacts with the ETS domain and the CID. Lysine acetylation of the NID relieves autoinhibition, likely through disruption of these intramolecular interactions. Mutational analyses revealed specific intramolecular linkages among the regulatory elements. From these findings, we propose a model for autoinhibition in the ETV1/4/5 subfamily in which structured and disordered regions regulate the DNA-recognition helix.

MATERIALS AND METHODS

Expression plasmids

Human *ETV1*, *ETV4*, *ETV5*, *ERG* and *FLI1* cDNAs corresponding to full-length or truncated proteins were cloned into the bacterial expression vector pET28 (Novagen) using standard sequence and ligation independent cloning strategies (25). Point mutations were introduced into the ETV4 plasmid using the QuikChange site-directed mutagenesis protocol (Stratagene). For acetylation studies, codons encoding Lys226 or Lys260 in the full-length *ETV4* gene were mutated to an amber codon (UAG), and the natural amber stop codon was mutated to an opal codon (UGA). Mutated *ETV4* cDNA was then cloned from the pET28 plasmid into a pCDF plasmid (kind gift from Dr. Jason Chin) for expression (26).

Expression and purification of proteins

All proteins were produced in *Escherichia coli* (λDE3) cells. Uninhibited ETS factor DNA-binding domains (DBDs) and the ETV1/4/5 fragments not containing the NID were efficiently expressed into the soluble fraction. Cultures of 1 L Luria broth (LB) were grown at 37°C to OD_{600} ∼ 0.7–0.9, induced with 1 mM isopropyl-β -D-thiogalactopyranoside (IPTG) and grown at 30°C for ∼ 3 h. To produce isotopically enriched proteins, expression was carried out using M9 minimal media supplemented with 3 g/l ($^{13}C_6$, 99%)-D-glucose and/or 1 g/l (^{15}N, 99%)-NH_4Cl.

Harvested cells were resuspended in 25 mM Tris pH 7.9, 1 M NaCl, 5 mM imidazole, 0.1 mM ethylenediaminetetraacetic acid (EDTA), 2 mM 2-mercaptoethanol (BME) and 1 mM phenylmethanesulfonylfluoride (PMSF). Cells were lysed by sonication and centrifuged at 125,000 × *g* for at least 30 min at 4°C. After centrifugation, the soluble supernatants were loaded onto a Ni^{2+} affinity column (GE Biosciences) and eluted over a 5–500 mM imidazole gradient. Fractions containing purified protein were pooled, combined with ∼ 1 U thrombin/mg of purified protein and dialyzed overnight at 4°C into 25 mM Tris pH 7.9, 10% glycerol (v:v), 1 mM EDTA, 50 mM KCl and 1 mM dithiothreitol (DTT). After centrifugation at 125,000 × *g* and 4°C, the soluble fraction was loaded onto a SP-sepharose cation exchange column (GE Biosciences) and eluted over a 50–1000 mM KCl gradient. Fractions containing the ETS proteins were loaded onto a Superdex 75 gel filtration column (GE Biosciences) in 25 mM Tris pH 7.9, 10% glycerol (v:v), 1 mM EDTA, 300 mM KCl and 1 mM DTT. Eluted fractions were analyzed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE). The final, purified proteins were then concentrated on a 10-kDa molecular weight cut-off (MWCO) Centricon device, snap-frozen with liquid nitrogen, and stored at -80°C in single-use aliquots for subsequent EMSA studies.

Full-length ETS factors and ETV4 truncations containing the NID generally expressed more efficiently in the insoluble fraction using an autoinduction protocol (27). Briefly, bacteria in 250 mL of autoinduction media were grown in 4 L flasks at 37°C to an OD_{600} ∼ 0.6 – 1. The temperature was then reduced to 30°C and cultures were grown for another ∼ 12 – 24 hr. Final OD_{600} values were typically ∼ 6 – 12, indicating robust autoinduction. Harvested cells were resuspended as described above, sonicated and centrifuged at 31,000 x *g* for 15 min at 4°C. The soluble fraction was discarded and this procedure was repeated with the pellet / insoluble fraction twice more to rinse the inclusion bodies. The final insoluble fraction was resuspended with 25 mM Tris pH 7.9, 1 M NaCl, 0.1 mM EDTA, 5 mM imidazole, 2 mM BME, 1 mM PMSF, and 6 M urea. After sonication and incubation for ∼ 1 hr at 4°C, the sample

was centrifuged at 125,000 x *g* for at least 30 min at 4°C. The soluble fraction was loaded onto a Ni^{2+} affinity column (GE Biosciences) and refolded by immediately switching to a buffer with the same components as above, except lacking urea. After elution with 5 – 500 mM imidazole, the remaining purification steps using ion-exchange and size-exclusion chromatography were performed as described above. However, a Q-sepharose anion-exchange column was used instead of a SP-sepharose cation-exchange column due to differing isoelectric points of the desired proteins.

Acetyllysine was incorporated into defined locations in the amino acid sequence and acetylated full-length ETV4 proteins were expressed according to a published protocol (26). Briefly, expression was induced with IPTG, as described above, but in the presence of 10 mM acetyllysine, 20 mM nicotinamide, and a plasmid expressing an amber tRNA that has been mutated to recognize acetyllysine. Acetylated proteins were purified as outlined above for unacetylated full-length ETV4.

ETV4 proteins prepared for NMR spectroscopy were purified using protocols slightly different from above. Harvested cells were resuspended in 50 mM sodium phosphate, 500 mM NaCl, 10 mM imidazole, 6 M guanidinium HCl, pH 7.4 and lysed by at least one round of freeze/thaw, followed by passage 5 times through an EmulsiFlex-C5 homogenizer at 10 kPa, and finally, 15 min of sonication. The cell lysate was spun down by centrifuging at 25,000 x *g* for 1 hr at 4°C. The supernatant containing ETV4 was then loaded onto Ni^{2+} affinity column (GE Biosciences), washed with 30 mM imidazole and eluted with 1000 mM imidazole and 6 M guanidinium HCl. Eluted fractions containing the desired protein were dialyzed against 3 L of refolding buffer (50 mM sodium phosphate, 1 M NaCl, 2 mM DTT and 1 mM EDTA, pH 7.5) at 4°C overnight. The His_6-tag of the refolded proteins was cleaved by adding 1 U of thrombin/mg or tobacco etch virus protease (TEV) at a TEV/protein ratio of 1/200 (w/w). The mixture was loaded onto another Ni^{2+} affinity column, and the flow-through containing the tag-free ETV4 fragment was concentrated using a 3 kDa MWCO Centricon device to 2 mL. Size exclusion chromatography with Superdex 75 was used for a last purification step. Eluted fractions were assessed using SDS-PAGE and those containing the purified protein were pooled and dialyzed against NMR sample buffer (20 mM sodium phosphate, 200–1000 mM NaCl, 2 mM DTT, 0.1 mM EDTA, pH 6.5).

Protein concentrations were determined by measuring the absorbance at 280 nm using predicted ϵ_{280} values, or at 595 nm after mixing 20 μl of protein with 1 ml of Bio-Rad Protein Assay Dye Reagent (diluted 1:5 in deionized water) and comparing to a bovine serum albumin standard curve. Molecular weights for each ETS protein were predicted using the Peptide Property Calculator (Northwestern University) or using the ExPASY web server (28).

Expressed protein ligation and purification

The DNA encoding ETV4 ETS domain and CID (residues 337–436) was sub-cloned into bacterial expression vector pEM5B (kind gift from Dr. Pierre Barraud, Université Paris Descartes) between XhoI and BamHI restriction sites. This enabled the addition of the required cysteine and TEV cleavage site (ENLYFQC) preceding the ETS domain, as described for the segmental labeling and expressed protein ligation protocol (29). The protein construct was expressed in LB media (unlabeled) or M9 media (^{15}N-labeled), purified under denaturing conditions, and refolded as described above. The protein was concentrated to 0.3 mM and stored in the inactive reaction buffer (50 mM HEPES, 200 mM NaCl, 0.1 mM TCEP, pH 7).

The DNA encoding ETV4 NID (165–336) was subcloned into pEM9B (kind gift from Dr. Pierre Barraud) between NdeI and SapI restriction sites. The pEM9B expression vector also encodes a C-terminal *Mxe* GyrA intein. Nine additional amino acids (GGGHM preceding and GSSC following the NID) were introduced as a result of cloning and to enable protein ligation. The protein construct was expressed in LB media (unlabeled) or M9 media (^{15}N-labeled), cells were harvested and resuspended in native buffer (50 mM sodium phosphate, 500 mM NaCl, 10 mM imidazole, pH 7.4) and lysed by cell homogenization and sonication, as described above. The supernatant containing the desired protein was purified first by loading onto the Ni^{2+} affinity column, washed by 30 mM imidazole and eluted with 1000 mM imidazole. The protein was concentrated to 0.5 mM and stored in the inactive reaction buffer (50 mM HEPES, 200 mM NaCl, 0.1 mM TCEP, pH 7).

Purified protein samples containing the ^{15}N-labeled ETV4 ETS domain and CID and unlabeled ETV4 NID were mixed in a 1:2 molar ratio. The reaction was activated by adding 100 mM 2-mercaptoethanesulfonate (MESNA) and TEV protease at a TEV/protein ratio of 1/200 (w/w). The reaction mixture was incubated at 16°C for 5 days. Time points were collected and analyzed on SDS-PAGE to monitor the ligation efficiency. TEV protease-cleaved products and intein self-cleaved products were purified on a chitin column equilibrated with 50 mM HEPES, 200 mM NaCl, pH 7. The flow-through of the chitin column containing the ligated product was purified on either ion-exchange chromatography (Mono Q) equilibrated with 50 mM HEPES pH 7 and eluted with 0–1000 mM NaCl gradient, and/or size exclusion chromatography (Superdex 75) equilibrated with NMR sample buffer (20 mM sodium phosphate, 200 mM NaCl, 2 mM DTT, 0.1 mM EDTA, pH 6.5). Fractions containing the final product were verified by SDS-PAGE and MALDI-ToF mass spectrometry on a Voyager-DE STR (Applied Biosystems) with a sinapinic acid matrix. The final product was dialyzed against NMR buffer. For the ligation reaction using ^{15}N-labeled ETV4 NID and unlabeled ETV4 ETS domain and CID, equal molar ratio were mixed (100 μM) to minimize aggregation due to highly concentrated ETV4 (337–436). The reaction was initiated and the final product was purified and confirmed, as described above.

Segmental isotope labeling using sortase A

The DNA encoding the Sortase A peptidase was a kind gift from Dr. Michael Sattler (Institute of Structural Biology, Helmholtz Zentrum München). A mutation of Gly to Ala immediately after the TEV cleavage site was made to optimize ligation efficiency. The plasmids encoding Sortase A

and ETV4 (337-436) were transformed into *E. coli* (λDE3) cells for protein expression at 37°C with 1 mM IPTG induction. Sortase A was purified by Ni^{+2} affinity and size exclusion chromatography. ^{15}N-labeled ETV4 (337–436) with an N-terminal glycine was prepared as described above, dialyzed into 50 mM Tris, 150 mM NaCl, pH 8.0, concentrated to 0.1 mM and stored at −80°C. ETV4 (165–336) was sub-cloned into pET28a between NdeI and XhoI restriction sites, with modification in the C-terminal to include the Sortase recognition sequence LPQTG plus a C-terminal His_6-tag. ETV4 (165–336) was expressed and purified using the same protocol as for ETV4 (337–436), except without thrombin digestion. Both Sortase A and ETV4 (165–336) were dialyzed into 50 mM Tris, 150 mM NaCl, pH 8.0 and concentrated to 0.5 mM for storage at −80°C. The ligation reaction was carried out as previously described (30). Briefly, the ^{15}N-labeled ETV4 fragment, the unlabeled ETV4 fragment and Sortase A were combined in a 2:6:1 molar ratio. The mixture was centifuged at 2000 x *g* in a 3 kDa MWCO Centricon device with reaction buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 10 mM $CaCl_2$) at 20°C for 4 h. The ligated product was purified by passing twice through a Ni^{+2} affinity column with thrombin cleavage after the first purification step, as described above. The final Sortase ligated sample of unlabeled ETV4 (165–336) linked to ^{15}N-labeled ETV4 (337–436) was verified by MALDI-ToF mass spectrometry and SDS-PAGE.

Electrophoretic mobility shift assays (EMSA)

DNA-binding assays of ETS factors utilized a duplexed 27-bp oligonucleotide with a consensus ETS binding site: 5′-TCGACGGCCAAGCC**GGAA**GTGAGTGCC-3′ (arbitrarily assigned as 'top' strand) and 5′-TCGAGGCACTCACTTCCGGCTTGGCCG-3′ ('bottom' strand). Boldface **GGAA** indicates the core ETS binding site motif. Each of these oligonucleotides, at 2 μM as measured by absorbance at 260 nM on a NanoDrop 1000 (Thermo Scientific), were labeled with [γ-^{32}P] ATP using T4 polynucleotide kinase at 37°C for ∼30–60 min. After purification over a Bio-Spin 6 chromatography column (Bio-Rad), the oligonucleotides were incubated at 100°C for ∼5 min, and then cooled to room temperature over 1–2 h. The DNA for EMSAs was diluted to 1×10^{-12} M and held constant, whereas protein concentrations ranged approximately six orders of magnitude. For full binding isotherms, the exact concentration range was chosen according to the K_D of particular protein fragments. Protein concentrations were determined after thawing each aliquot of protein, using the Protein Assay Dye Reagent. Equivalent starting amounts (0.2 μg) of each protein utilized on a given day were run on an SDS-PAGE gel to confirm their relative concentrations. The binding reactions were incubated for 45 min at room temperature in a buffer containing 25 mM Tris pH 7.9, 0.1 mM EDTA, 60 mM KCl, 6 mM $MgCl_2$, 200 μg/ml BSA, 10 mM DTT, 2.5 ng/μl poly(dIdC) and 10% (v:v) glycerol, and then resolved on an 8% (w:v) native polyacrylamide gel at room temperature. The ^{32}P-labeled DNA was quantified on dried gels by phosphorimaging on a Typhoon Trio Variable Mode Imager (Amersham Biosciences). Equilibrium dissociation constants (K_D) were determined by nonlinear least squares fitting of the total protein concentration $[P]_t$ at each titration point versus the fraction of DNA bound ($[PD]/[D]_t$) to the equation $[PD]/[D]_t = 1/[1 + K_D/[P]_t)]$ using Kaleidagraph (v. 3.51; Synergy Software). The DNA, running at the position of free DNA, is unbound and by comparing the DNA signal in this position, [D], to the DNA signal in lanes with no binding at the lowest protein concentrations, $[D]_t$, we calculated the fraction of bound DNA, $[PD]/[D]_t = 1 - [D]/[D]_t$. Due to the low concentration of total DNA, $[D]_t$, in all reactions relative to the K_D, the total protein concentration is a valid approximation of the free, unbound protein concentration $[P] = [P]_t$. Reported K_D values represent the mean of at least three independent experiments and the standard error of the mean. We used Microsoft Excel (v.14.6.5 for Mac 2011) to perform two-tailed, heteroscedastic *t*-tests in order to compare the K_D values of different proteins. This method provided the same *P*-values as using a Welch's *t*-test.

Although there were deviations between the EMSA titration data and the non-linear regression at higher protein concentrations, this had essentially no influence on calculated K_D values. For comparison we fit all data using a one-site specific binding model with a Hill coefficient (Prism, version 6), while treating the maximum fraction of DNA bound as a variable. Use of this alternative approach, which eliminated the deviation between the data and the curve fit at the higher protein concentrations, led to K_D values that were uniformly stronger by about 1.3–1.7-fold, depending on the protein tested, with Hill coefficients ranging from 0.9 to 1.1. Importantly, all fold-inhibition values changed <1.5-fold, and all conclusions made from the binding experiments were supported by either curve fitting approach.

To test protein activity, binding reactions with known concentrations of radiolabeled duplex DNA (as described above) were titrated against a fixed concentration of protein corresponding to ∼ 50-fold greater than the K_D value of each individual protein, as previously described (31). All proteins analyzed demonstrated high activity levels (>95 %), indicating that the autoinhibition of DNA binding observed for the larger protein species is not due to a large fraction of the protein being totally inactive in DNA binding.

Partial proteolysis

For tryptic digestion studies, 20 μl of full length (FL) ETV4 at 20 μM was incubated with 1.5–450 ng of trypsin (Sigma) in a buffer containing 25 mM Tris pH 7.9, 10 mM $CaCl_2$, and 1 mM DTT. After 2 min of incubation, the reaction was quenched with 1 % (v:v) acetic acid (final volume). The resulting samples were analyzed by SDS-PAGE and ESI-MS (total mixture analyzed), and used for EMSA studies.

Crystallization and structure determination

Purified proteins were dialyzed overnight in 10 mM Tris pH 7.9 and 50 mM NaCl, and then concentrated to 5 mg/ml. Crystals were grown by vapor diffusion in sitting drops of 2:1 protein:reservoir (v:v). CID-inhibited ETV1 (332–435) was crystallized against a reservoir of 30% (w:v) PEG 5000

monomethyl ether, 0.1 M MES sodium salt and 0.2 M ammonium sulfate at pH 6.5 and 20°C. CID-inhibited ETV4 (337–441) was crystallized against a reservoir of 1 M diammonium phosphate and 0.1 M sodium acetate at pH 4.5 and 20°C. Uninhibited ETV5 (364–457) was crystallized against a reservoir of 0.2 M diammonium phosphate and 20% PEG 3350 at pH 5.0 and 4°C.

Crystals were immersed briefly in mother liquor containing 20% glycerol, and then cryocooled by plunging into liquid nitrogen. Diffraction data were collected on a Q315 CCD using Stanford Synchrotron Radiation Lightsource (SSRL) beamline 7-1 with X-rays at 1.0000 Å (ETV1 and ETV4) or 1.1271 Å (ETV5). The resulting data were integrated and scaled using HKL2000 (32). Phases were determined by molecular replacement with Phaser-MR (33) using the ETS domain of ETS1 (1MD0.pdb) as a search model. Models were built with COOT (34) and refined with PHENIX (35). PyMOL (Schrödinger, LLC) was used to render molecular structure figures.

Model geometries were analyzed by MolProbity (36) within PHENIX. For ETV1 (1.4 Å resolution data), 87.5% of residues have favorable backbone dihedrals and 12.5% fall into allowed regions. Residues 332–333 and 435 were not visible in the electron density. For ETV4 (1.1 Å), 91.8% of residues have favorable backbone dihedrals and 8.2% of residues fall into allowed regions. Residues 337–339 and 337–441 were not visible in the electron density. For ETV5 (1.8 Å), 87.7% of residues have favorable backbone dihedrals and 12.3% of residues fall into allowed regions. Residues 364–365 were not visible in the electron density. X-ray crystallography data collection and refinement statistics are provided in Supplementary Table S1. Structural coordinates have been deposited in the RCSB Protein Data Bank under ID codes 5ILS (ETV1), 5ILU (ETV4) and 5ILV (ETV5).

Circular dichroism spectroscopy

Aliquots of frozen ETV4 NID (165–336), expressed and purified as described above, were thawed, dialyzed overnight into 20 mM sodium phosphate, 50 mM NaCl, pH 7.9, and diluted to 25 μM concentration. CD spectra were recorded at 4°C through the wavelength range of 190–260 nm with a 1 nm wavelength step. A baseline reference, consisting of buffer only, was subtracted from the CD spectra. Three scans were collected in series and averaged after visually verifying their consistency. Data were converted to molar ellipticity as described (37).

NMR spectroscopy

NMR data were recorded at 25°C on cryoprobe-equipped 500, 600 and 850 MHz Bruker Avance III spectrometers. Proteins were in NMR sample buffer (plus 10% lock D_2O) with 1 M NaCl for spectral assignments and with 200 mM NaCl for all other experiments. The elevated ionic strength reduced slow aggregation over long-term measurements. Data were processed and analyzed using NMRpipe (38) and Sparky (39). Signals from mainchain and sidechain ^{1}H, ^{13}C and ^{15}N nuclei were assigned by standard multidimensional heteronuclear correlation experiments, including ^{15}N-HSQC, HNCO, HN(CA)CO, CBCA(CO)NH and HNCACB (40). Amide $^{1}H/^{2}H$ hydrogen exchange (HX), after transfer into ∼99% D_2O NMR sample buffer via a Sephadex G25 spin column and CLEANEX-PM $^{1}H/^{1}H$ HX measurements were recorded using 850 MHz NMR spectrometer and analyzed as described previously (10,41). The two approaches detect slow (minutes-days) and fast (seconds) timescale exchange, respectively. Initially, HX measurements were carried out at pH 6.5 and 25°C with uninhibited ETV4 (328–430). However, a more complete set of data were obtained for CID-inhibited ETV4 (337–436) via $^{1}H/^{2}H$ exchange measurements at pH 5.7 and 20°C and CLEANEX-PM $^{1}H/^{1}H$ exchange at pH 7.5 and 25°C. Protection factors were calculated as ratio of the predicted HX rate constant for each amide in an unstructured polypeptide with the sequence of ETV4 versus the corresponding measured HX rate constant. The predicted values, corrected for pH, temperature and isotope effects, were obtained with the program Sphere (http://landing.foxchase.org/research/labs/roder/sphere/) (42). The protection factors for the two proteins, studied under several conditions, were merged and reported using the combined name ETV4 (328–436) (Figure 5C). For amides with HX quantitated under more than one condition, the highest protection factor is shown.

RESULTS

DNA binding by ETV1, ETV4 and ETV5 is autoinhibited

We initially sought to determine the magnitude of autoinhibition in the ERG and ETV1/4/5 subfamilies of ETS factors. Toward this aim, we measured the DNA binding affinities (equilibrium dissociation constant, K_D) of the full-length proteins and nearly-minimal DBDs for ERG, FLI1, ETV1, ETV4 and ETV5 (Figure 1A). Robust autoinhibition was observed in ETV1, ETV4, and ETV5, with the full-length proteins displaying ∼10 to 30-fold weaker binding than their minimal DBDs (Figure 1B–D and Supplementary Table S2). These levels of autoinhibition are comparable to those previously reported for ETS1 (8) and ETV6 (9). In contrast, ERG and its subfamily member FLI1 displayed modest 2- to 3-fold autoinhibition, as also previously reported (43). Interestingly, the K_D values cluster in a pattern that reflects their subfamily phylogenetic classifications (Figure 1E) (44). The ∼100-fold range of K_D values for ETS DBDs suggests that there are key differences that influence DNA binding despite the high overall sequence conservation in the ETS domain of these factors. Additionally, the known inhibitory elements from ETS1 and ETV6 are not conserved in the ETV1/4/5 or ERG subfamilies; the poor sequence conservation outside of the ETS domain among these factors indicates that the mechanism of autoinhibition is likely different for the ETV1/4/5 and ERG subfamilies (Supplementary Figure S1). Based on the larger magnitude of autoinhibition observed with ETV1, ETV4, and ETV5, as compared to ERG and FLI1, we focused on the ETV1/4/5 subfamily for mechanistic studies.

We chose ETV4 as a model factor to further investigate autoinhibition in the ETV1/4/5 subfamily. Partial proteolysis aided the design of truncation boundaries for mapping inhibitory elements (Supplementary Figure S2A–C). We found that the predominant trypsin-resistant fragment,

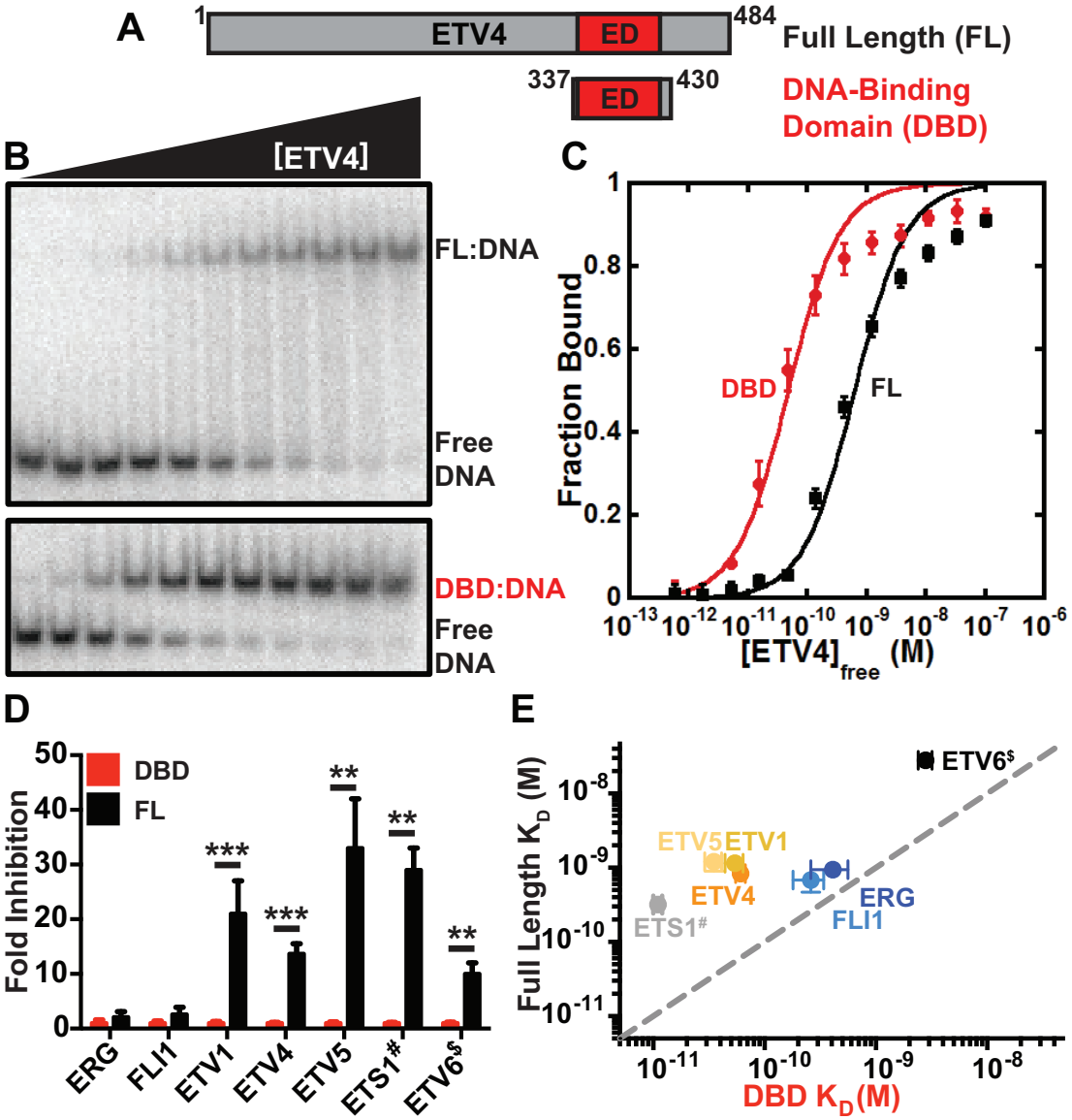


Figure 1. Autoinhibition in the ERG and ETV1/4/5 subfamilies. (**A**) Schematic of full-length protein, FL, and nearly minimal DNA-binding domain, DBD, for ETV4. Based on the sequences of all ETS factors, the conserved ETS domain, ED, is noted in red. (**B**) Representative examples of EMSA gels for ETV4 FL or DBD with a double-stranded DNA duplex containing a core ETS binding site. (**C**) Binding isotherms for ETV4 FL and DBD. Data points and error bars refer to the mean and the standard error of the mean with four replicates for each protein. See 'Materials and Methods' section for details. (**D**) Fold inhibition of ERG, FLI1, ETV1, ETV4 and ETV5, calculated as K_D (FL or DBD) / K_D (DBD). ETS1 (8)# and ETV6 (9)$ data are included for comparison. Mean and standard error of the mean from at least three replicates are plotted; '**' and '***' indicate $P < 0.01$ and $P < 0.001$, respectively. See Supplementary Table S2 for K_D values and numbers of replicates. (**E**) K_D values of FL versus DBD for each of the ETS factors tested. The dashed diagonal line represents no autoinhibition [i.e. K_D (FL) = K_D (DBD)].

spanning amino acids 165–484, retained levels of autoinhibition comparable to full-length ETV4 (Figure 2A and Supplementary Table S3). Subsequent deletion studies revealed that amino acid residues both N- and C-terminal of the ETS domain inhibit DNA binding independently, but also act cooperatively to yield higher than additive levels of inhibition (Figure 2B). Hereafter, these regions will be denoted as the NID and the CID, whereas the nearly-minimal DBD will be denoted as an uninhibited species. We hypothesized that the ETV1/4/5 NID and CID function through direct interactions with the ETS domain and/or with each other to cooperatively inhibit DNA binding (Figure 2C).

CID interactions perturb the DNA-recognition helix H3 to mediate autoinhibition

To elucidate the mechanism of autoinhibition by the CID, we used X-ray crystallography to determine the structures of the partially inhibited fragments of ETV1 (332–435) and ETV4 (337–441). These proteins contain both the ETS domain and the CID, as mapped in ETV4. Their structures were very similar with a root mean square deviation

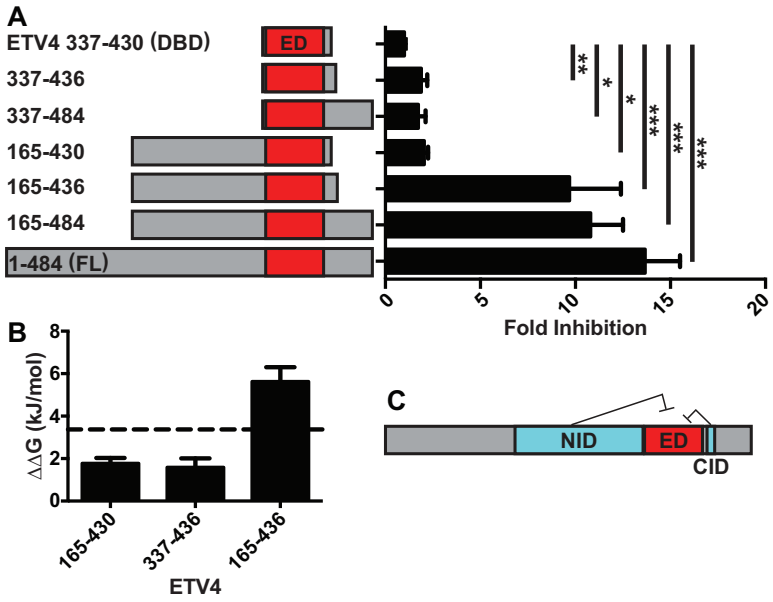


Figure 2. NID and CID cooperate to inhibit ETV4 DNA binding. (**A**) Fold inhibition of the ETV4 fragments with mean and standard error of the mean displayed. Fold inhibition calculated as K_D (fragment) / K_D (337–430). '*', '**', And '***' indicate $P < 0.05$, $P < 0.01$ and $P < 0.001$, respectively. See Supplementary Table S3 for K_D values and numbers of replicates. (**B**) $\Delta\Delta G = RT \ln [K_D$ ETV4 inhibited fragment / K_D ETV4 (337–430)] measured for fragments containing the NID (165–430), the CID (337–436) or both (165–436). The dotted line indicates the sum of the $\Delta\Delta G$ values for 165–430 and 337–436. (**C**) Schematic of ETV4 autoinhibition depicting cooperative inhibitory contributions from both the NID and CID, cyan. The ETS domain (ED) is noted in red.

(RMSD) of 0.16 Å for backbone alignment of their ETS domains (Figure 3A and B; Supplementary Figure S3A). The CID includes an α-helix, termed H4, which packs on one face of the ETS domain. In ETV4, Ala426 and Leu430 in H4 lie in a hydrophobic groove along the ETS domain in proximity to the conserved residues Trp344 from H1, Ile407 from the loop between H3 and β-strand S3, and Phe420 from S4 (Figure 3C). Homologous residues had similar interactions in ETV1. Replacing Leu430 with an alanine resulted in a reduction in autoinhibition (activation in DNA binding), whereas mutation to methionine, the homologous amino acid in ETV1 and ETV5, did not affect DNA binding (Figure 3D). These structural and functional data demonstrated that the CID inhibits DNA binding through intramolecular contacts between H4 and the ETS domain, mediated in part by a leucine or methionine in this helix.

Based on the crystal structures of CID-inhibited ETV1 and ETV4, we noted that the uninhibited, minimal DBD fragments used for demonstrating autoinhibition in ETV1, ETV4 and ETV5 were predicted to have a shorter or possibly unfolded helix H4 (Supplementary Figure S3B). As with ETV4, loss of these homologous residues in ETV1 and ETV5 also activated DNA binding (Figure 4A). Therefore, an intact and full-length H4 is a necessary and conserved feature of the CID.

To understand the structural nature of the residues mapped to H4 within the context of uninhibited ETV1, ETV4 and ETV5, we attempted to crystalize these fragments with success for ETV5 (364–457) (Figure 4B). Despite the deletion of amino acids mapped to the intact H4, the α-helix is retained, albeit truncated. However, the shorter H4 is rotated ∼60° away from the ETS domain relative to the position of the full-length H4 in ETV1 and ETV4. This alternate position is accommodated in the crystal by intermolecular contacts between the truncated H4 and the ETS domain of a neighboring molecule (Supplementary Figure S4). With H4 in this alternate position, Met457 is unable to form the intramolecular inhibitory contacts with the ETS domain observed for the homologous Met424 and Leu430 in the CID-inhibited structures of ETV1 and ETV4, respectively, potentially explaining the loss of autoinhibition of this fragment (compare Figures 3B and 4B). In conclusion, the relief of autoinhibition by the partial truncation of H4 and by disruption of an intramolecular contact between H4 and the ETS domain demonstrated the role of H4 in autoinhibition. In addition, while the folding of truncated helix H4 and its alternate position is potentially a consequence of crystallization, we propose that this repositioning reflects an intrinsic mobility of the CID. This idea is supported by NMR spectroscopy studies, presented below.

To further our structural studies of ETV1/4/5, we compared our crystal structures of the uninhibited ETV5 with a truncated H4 to that of the CID-inhibited ETV1 and ETV4 with a full-length H4. In comparison to the highly similar CID-inhibited ETV1 and ETV4 structures (backbone RMSD of 0.16 Å), the ETS domain from uninhibited ETV5 was distinct with RMSD values of 0.79 Å and 0.72 Å when aligned to ETV1 and ETV4, respectively (Supplementary Figure S3A). Closer examination of subsections of the ETS domain revealed that the differences between uninhibited and CID-inhibited structures were most pronounced around the DNA-recognition helix H3, as well as β-strands S3 and S4. Visually, the backbone of the C-terminal half of the DNA-recognition helix H3 is shifted about 2 Å between the inhibited and uninhibited structures, relative to the rest of the ETS domain (Figure 4C). Further comparison with

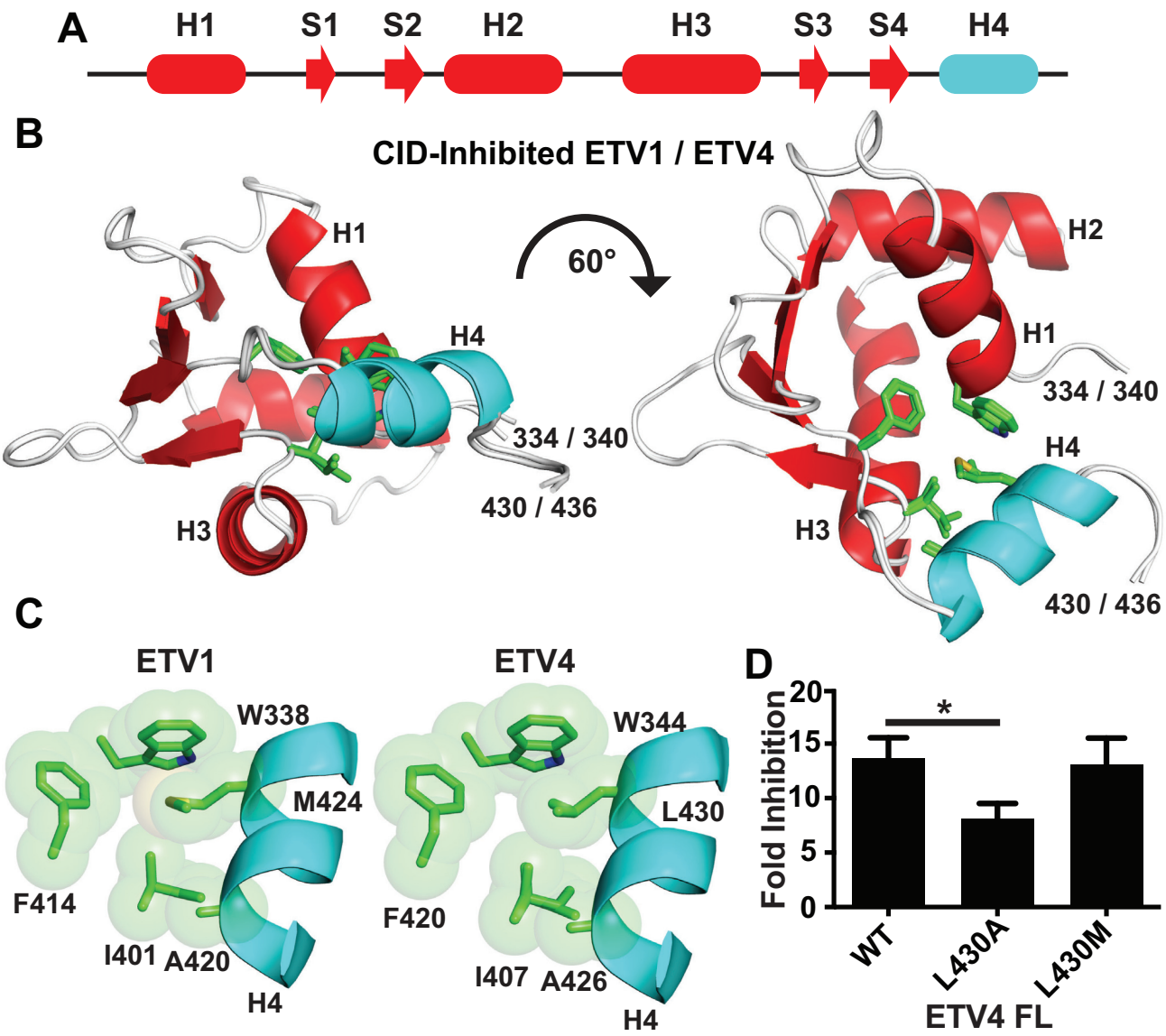


Figure 3. The CID inhibits DNA binding through hydrophobic contacts between α-helix H4 and the ETS domain. (**A**) Schematic of ETS domain, H1-H3 and S1-S4, and α-helix H4 of ETV1, ETV4 and ETV5. ETS domain, red; inhibitory elements, cyan; α-helices, cylinders; β-strands, arrows. (**B**) Cartoon representations of the closely aligned structures for the ETS domain and CID of ETV1 (332–435) and ETV4 (337–441). Displayed in stick format are Ala426 and Leu430 from α-helix H4 in ETV4, and the analogous Ala420 and Met424 from ETV1, as well as the conserved amino acids in the ETS domain that form a hydrophobic cluster. Numbering for homologous amino acids and endpoints denoted as ETV1/ETV4. (**C**) Portions of the ETV1, left, and ETV4, right, structures in van der Waals sphere format to show hydrophobic interactions between the ETS domain and H4. There is clear evidence for two conformations of the δ^1-methyl of Ile407 in the electron density map of ETV4. (**D**) Fold inhibition of ETV4 FL in its wild-type form, WT ($n = 35$) or with point mutations L430A ($n = 11$) or L430M ($n = 3$). '*' Indicates $P < 0.05$.

the previously reported structure of ETV4 in complex with DNA (45) demonstrated that in the DNA-bound form, H3 of ETV4 is also shifted to a similar position as observed for uninhibited ETV5 (Figure 4C and Supplementary Figure S3C). We speculate that in the ETV1/4/5 subfamily, the active state of a DNA-bound ETS domain requires this shift of H3 and, thus, matches the conformation of uninhibited ETV5 determined by X-ray crystallography.

Having observed the activation of the ETV4 mutant L430A (Figure 3D) and the variable positioning of the DNA-recognition helix H3 in our crystal structures (Figure 4C), we hypothesized that helix H4 inhibits DNA binding by modulating H3 through an interaction between Leu430 in H4 and Ile407 in the H3-S3 loop. We tested this postulate by mutating Ile407 and Leu430 to alanine, separately and in combination. The ETV4 mutant I407A had a reduction in DNA-binding affinity compared to the wild-type protein indicating that I407A contributes to DNA binding. Importantly, the I407A mutation also abrogated the activating nature of L430A in the double mutant I407A/L430A (Figure 4D). We conclude that H3 and the CID are coupled through the Ile407-Leu430 interaction and propose that CID-mediated autoinhibition functions by shifting a

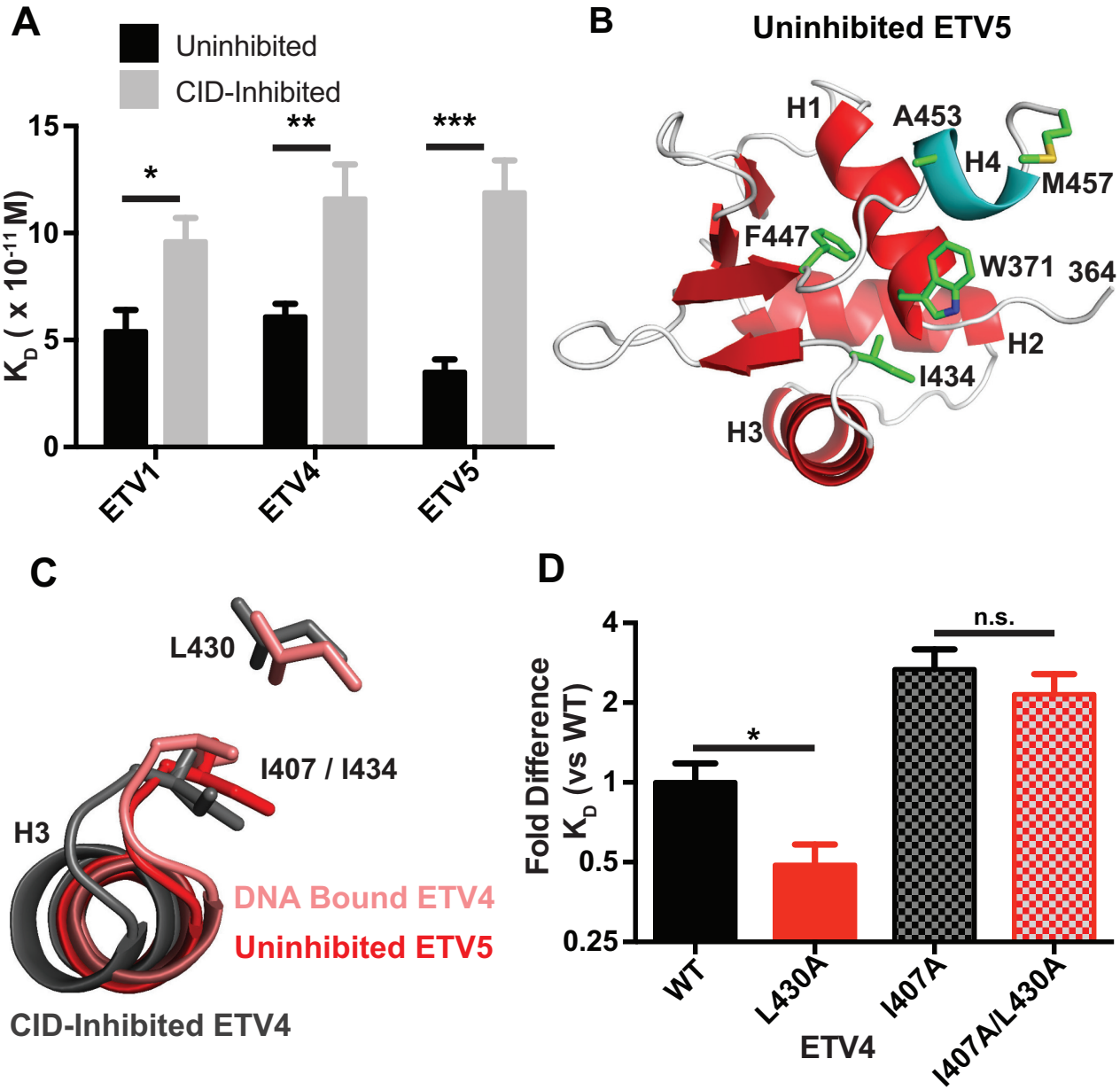


Figure 4. Interactions between the CID and the ETS domain affect DNA-recognition helix H3 positioning. (**A**) Equilibrium dissociation constant, K_D, values for uninhibited ETV1 (residues 332–425; $n = 6$), ETV4 (337–430; $n = 25$) and ETV5 (364–457; $n = 4$) versus CID-inhibited ETV1 (332–430; $n = 7$), ETV4 (337–436; $n = 23$) and ETV5 (364–463; $n = 7$). '*', '**' and '***' indicate $P < 0.05$, $P < 0.01$ and $P < 0.001$, respectively. (**B**) Crystal structure of uninhibited ETV5 (364–457), showing the truncated H4 and the same selected sidechains as in Figure 3. (**C**) H3 positioning from CID-inhibited ETV4 (gray), uninhibited ETV5 (red) and ETV4 bound to DNA (pink, 4UUV.pdb) (45). Structures were aligned to DNA-bound ETV4 across the entire protein sequence (See Supplementary Figure S3C). Met457 of ETV5, the homologous residue to Leu430 in ETV4, is not in frame due to the repositioning of H4 in the uninhibited ETV5 crystal structure. (**D**) Comparison of K_D values for ETV4 FL in its WT form (n = 35), or with point mutations L430A ($n = 11$), I407A ($n = 4$) or both I407A and L430A ($n = 4$). '*' Indicates $P < 0.05$ and 'n.s.' indicates $P > 0.05$. Fold difference for K_D values are relative to WT ETV4 FL.

conformational equilibrium of H3 toward a state that is less competent for DNA binding.

Dynamic features of CID autoinhibition mechanism

To further investigate the CID mechanism of autoinhibition, we utilized NMR spectroscopy to compare uninhibited and CID-inhibited species (Figure 5A). Although differing slightly at their N- and C-terminal boundaries, the two ETV4 fragments displayed the same affinities for DNA as the corresponding species discussed above (Supplementary Figure S5A). Based on mainchain chemical shifts, residues from the truncated H4 in the uninhibited ETV4 (328–430) and the full-length H4 in the CID-inhibited ETV4 (313–46) both adopted folded α-helical conformations under solution conditions. However, relative to ETV4 (313–446), the C-terminal residues in the shorter H4 of

ETV4 (328–430) exhibited reduced chemical shift-derived helical propensities and increased mobility as detected by amide ^{15}N relaxation measurements (Supplementary Figure S5B and data not shown). Nevertheless, these NMR spectroscopic studies indicate that the truncated α-helix H4 observed by X-ray crystallography is not an artefact of the crystallization process (Figure 4B).

A comparison of the $^{1}N^{H}$ and ^{15}N chemical shifts of the uninhibited and CID-inhibited species demonstrated that amides near the C-terminal end of H1 and H3, and throughout H4 were most affected by the presence of the full-length versus truncated H4 (Figure 5B). The amino acids in H3 that showed chemical shift perturbations match closely to those undergoing the backbone realignment observed in the comparison of the crystal structures of CID-inhibited ETV1 and ETV4 versus uninhibited ETV5. Thus, the interaction between H4 and the ETS domain, as well as the H4-dependent perturbations of H3, observed in the crystal structures are also retained in solution.

Additional NMR-monitored amide HX experiments were used to probe the dynamics of CID-inhibited ETV4 (Figure 5C). Residues within H1, H2 and the β-sheet displayed relatively large protection factors ($>10^4$), indicating that they form the stable core of the ETS domain. In contrast, residues preceding the ETS domain and in loop regions had lower protection factors, as expected based on their solvent exposure and lack of any persistent hydrogen-bonded secondary structure. Most interestingly, many residues in the DNA-recognition α-helix H3 and in the inhibitory CID α-helix H4, displayed intermediate protection factors (100–10,000) indicative of conformational dynamics to sample partially unfolded states detectable by HX. Similar behavior is observed with the DNA-recognition and inhibitory helices of ETS1 (8,12) and ETV6 (10,11). These NMR experiments extend our hypothesis that the CID autoinhibitory mechanism involves a conformational equilibrium involving interactions between helices H3 and H4, by demonstrating the dynamic nature of the helices.

Inhibitory properties of the NID map to intrinsically disordered sequences

As the next step toward a mechanistic understanding of ETV4 autoinhibition, we investigated the NID using biophysical approaches. Crystallization of fragments of ETV4 encompassing the NID was unsuccessful, potentially reflecting the predicted disordered nature of the NID (Supplementary Figure S2D). Consistent with this prediction, the ^{15}N-HSQC spectrum of the isolated ETV4 NID (165–336) displayed limited $^{1}H^{N}$ chemical shift dispersion (Figure 6A). An analysis of its assigned mainchain ^{1}H, ^{13}C and ^{15}N chemical shifts confirmed that the NID predominantly samples random coil conformations and thus is an IDR (Supplementary Figure S6) (46,47). Circular dichroism spectroscopy added additional evidence for the overall disordered character of the NID (Supplementary Figure S7).

Many IDRs, while disordered in isolation, take on a more structured character in the presence of a binding partner through a coupled 'folding and binding' mechanism (48). Therefore, we asked whether the NID is still disordered in the presence of the ETS domain and CID. Using intein and sortase ligation technologies, we covalently linked the ^{15}N-labeled NID (165–336) to unlabeled ETS domain and CID of ETV4 (337–436). We confirmed that this ligated fragment displayed comparable autoinhibition to the native protein fragment (Supplementary Figure S8). The NID spectrum retained limited $^{1}H^{N}$ chemical shift dispersion, indicating the lack of any detectable structure induced upon covalent-linkage to the ETS domain and CID (Figure 6A). Although small changes in the chemical shifts or relative intensities of the $^{1}H^{N}$-^{15}N signals from amides localized near the N-terminus of the NID were observed (Figure 6B and C), no segment of residues interacted with the ETS domain and/or CID of ETV4 with sufficient affinity to adopt a persistent conformation detectable by NMR spectroscopy. Thus, whether isolated or 'attached in cis', the NID was observed to be predominantly disordered.

We next interrogated which regions of the NID are important for inhibition. A truncation series indicated that the progressive inclusion of residues N-terminal to the ETS domain provided progressively greater autoinhibition of DNA binding (Supplementary Figure S9). The region spanning residues 203–287 contributed the largest effect of autoinhibition, but other regions of the NID also contributed toward the overall inhibitory effect. The lack of a clear boundary for the inhibitory residues is consistent with the intrinsic disorder of the NID and the absence of any identifiable cluster of residues that interact strongly with the ETS domain and CID.

Intramolecular interactions of NID with the ETS domain and CID

To better define the possible interactions of the NID with the ETS domain and CID, we also compared the ^{15}N-HSQC spectra of an ETV4 fragment containing the ETS domain and the CID only (337–436), with one also containing the NID (165–436) (Figure 7A). As described above, for spectral simplification, an unlabeled NID was joined by intein technology to a ^{15}N-labeled fragment containing the ETS domain and CID. In addition to the expected changes at the N-terminus of H1, the ligated NID weakly perturbed the $^{1}H^{N}$-^{15}N signals of amides in H2, the C-terminal region of H3 and the surface-exposed face of the CID (Figure 7B and C). These data suggest that the NID may inhibit DNA binding by transiently interacting with the DNA-recognition helix H3 and/or by interacting with, and reinforcing the inhibitory position of, the CID.

The largest chemical shift perturbations, besides those near the N-terminus ligation site, were observed for Tyr401 and Tyr403 on the DNA-recognition helix H3. Therefore, we tested whether these tyrosine residues are functionally important for autoinhibition by serving as an interaction site within the ETS domain for the NID. Tyr401 and Tyr403 were mutated to asparagine and glycine, respectively, in the CID-inhibited fragment of ETV4 (337–436). These alternative residues were chosen due to their presence in homologous positions in other ETS factors, suggesting that there would be less chance of structural perturbations. However, due to their position at the DNA interface, mutation of both

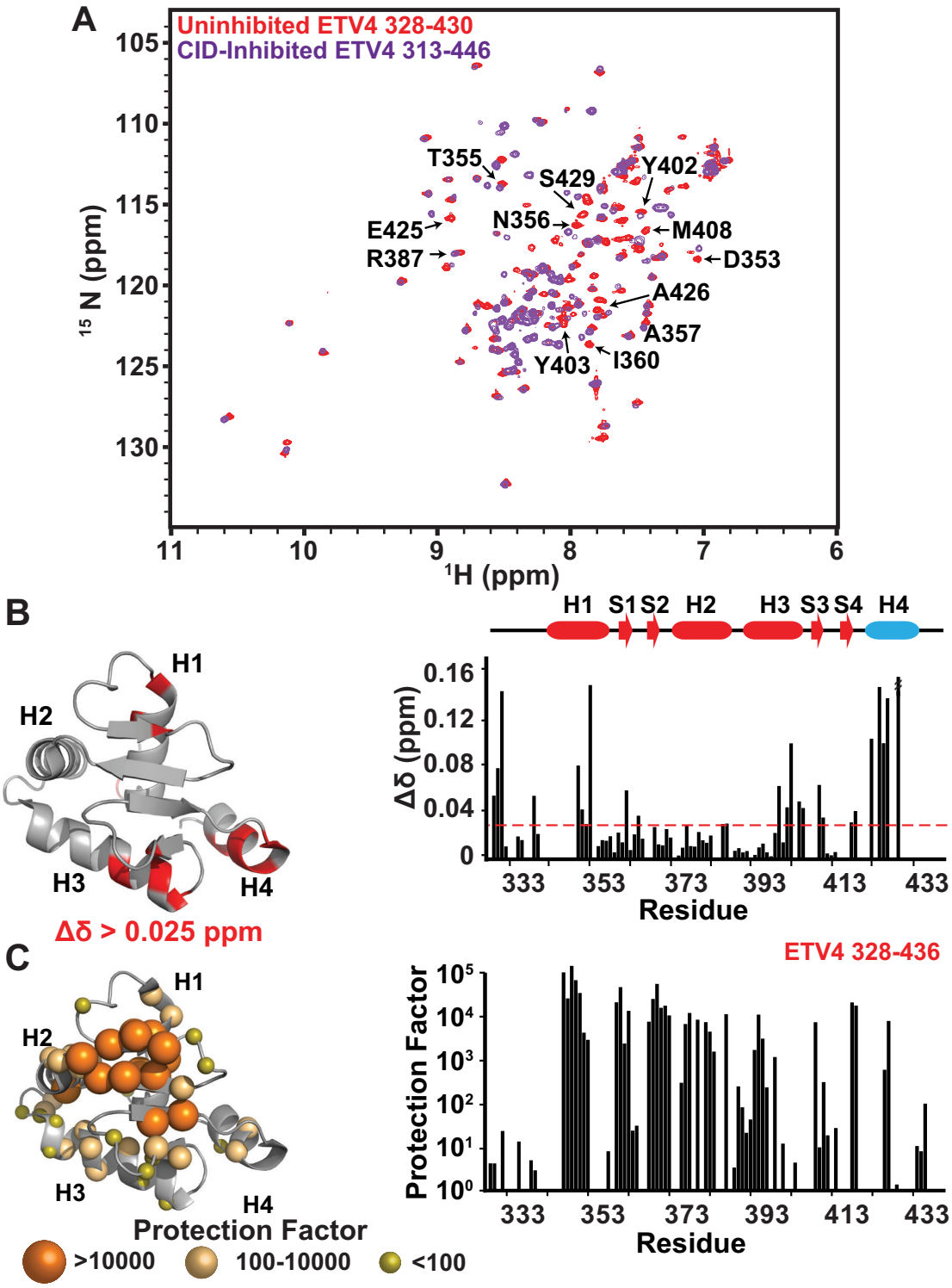


Figure 5. The CID perturbs the dynamic DNA-recognition helix H3. (**A**) Overlaid ^{15}N-HSQC spectra of uninhibited ETV4 (328–430; red), and CID-inhibited ETV4 (313–446; purple). Selected assignments are indicated. Despite minor differences in the boundaries of these constructs, they bind to DNA with similar affinities as the previously described uninhibited and CID-inhibited fragments (Supplementary Figure S5A). (**B**) Amide chemical shift perturbations ($\Delta\delta = [(\Delta\delta_H)^2 + (0.2\Delta\delta_N)^2]^{\frac{1}{2}}$) for corresponding residues in the spectra of (A) are plotted as a histogram and mapped onto the crystal structure of ETV4. Perturbed residues with $\Delta\delta > 0.025$ ppm (dashed line), are highlighted in red on the structure. (**C**) Amide HX protection factors of ETV4 (328–436), are plotted as a histogram and mapped onto the crystal structure of ETV4 using spheres with the indicated size/color scale. Although ETV4 (328–436) is a combined name to denote data merged from HX studies of uninhibited ETV4 (328–430) and CID-inhibited ETV4 (337–436), the additional six C-terminal residues, which form the full CID, did not measurably change the protection factors of amides within the ETS domain. Missing values correspond to prolines, residues with unassigned or overlapped NMR signals and residues exchanging too slowly to be measured with CLEANEX-PM (sec timescale), yet too fast to measured via ^{1}H/^{2}H exchange (greater than hours).

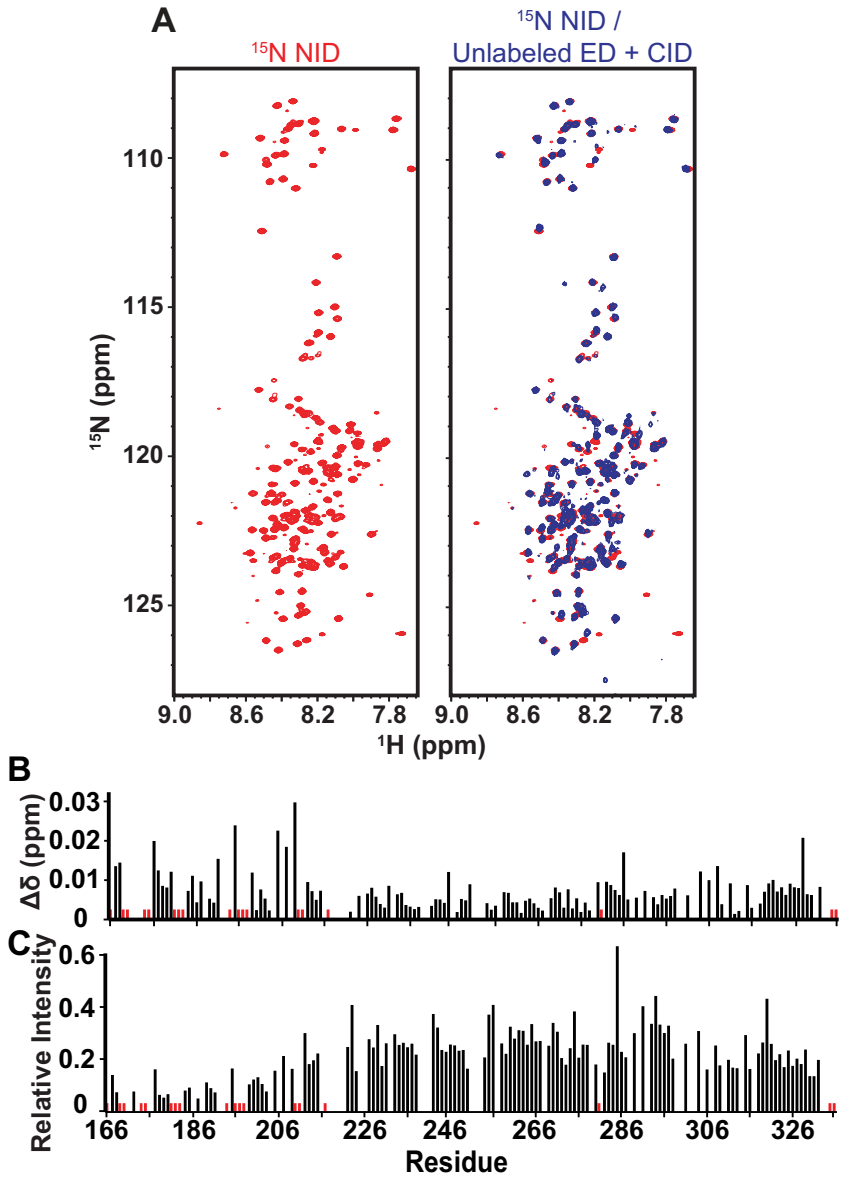


Figure 6. The NID is intrinsically disordered. (**A**) Both panels show the ^{15}N-HSQC spectrum of the isolated ETV4 NID (165–336) in red. The right panel also shows the overlapped spectrum of the ^{15}N-labled NID intein-ligated to the unlabeled ETS domain and CID (337-436) in blue. Consistent with its limited $^{1}H^{N}$ dispersion, analysis of the assigned mainchain chemical shifts ($^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, ^{13}CO) of the isolated NID with the algorithms CSI 2.0 (46) and δ2D (47) confirms that it is an IDR (Supplementary Figure S6). Comparisons of ^{15}N-HSQC amide chemical shift perturbations ($\Delta\delta = [(\Delta\delta_H)^2 + (0.2\Delta\delta_N)^2]^{\frac{1}{2}}$), (**B**), and relative peak intensities, (**C**), for the NID alone versus attached to the ETS domain and CID. The simple peak intensity ratios were not normalized for differences in sample concentration and spectral acquisition times. Amides broadly localized near the N-terminus of the NID showed small chemical shift and relative intensity perturbations due to the ETS domain and CID 'in cis'. The ^{15}N-HSQC spectrum of the intein-ligated species was assigned by comparison with that of the isolated NID, and red bars indicate amide signals that could not be confidently identified in both proteins. Missing histogram bars correspond to prolines and unassigned amides.

residues substantially impaired DNA binding (Supplementary Figure S10A). Importantly, full-length ETV4 and the CID-inhibited fragment of ETV4 with Y401N and Y403G substitutions had identical affinity for DNA, indicating a loss of NID-mediated inhibition (Figure 7D and Supplementary Figure S10A). These data indicate that Tyr401 and Tyr403 are required for NID-mediated inhibition and may serve as part of a transiently occupied interface. Due to the role of these two tyrosines in direct contact to DNA, this NID intramolecular interaction may be acting sterically by masking the DNA interface.

Another potential NID-interaction site mapped to the surface of the CID. However, mutation of several glutamate or phenylalanine residues along the surface-exposed face of the CID (E423K, E425K or F428A/F432A) did not affect NID-mediated inhibition (Supplementary Figure S10B). Therefore, two possibilities exist to explain the NMR spectroscopically detected perturbation of the CID by the NID. The potential NID–CID interface may be sufficiently

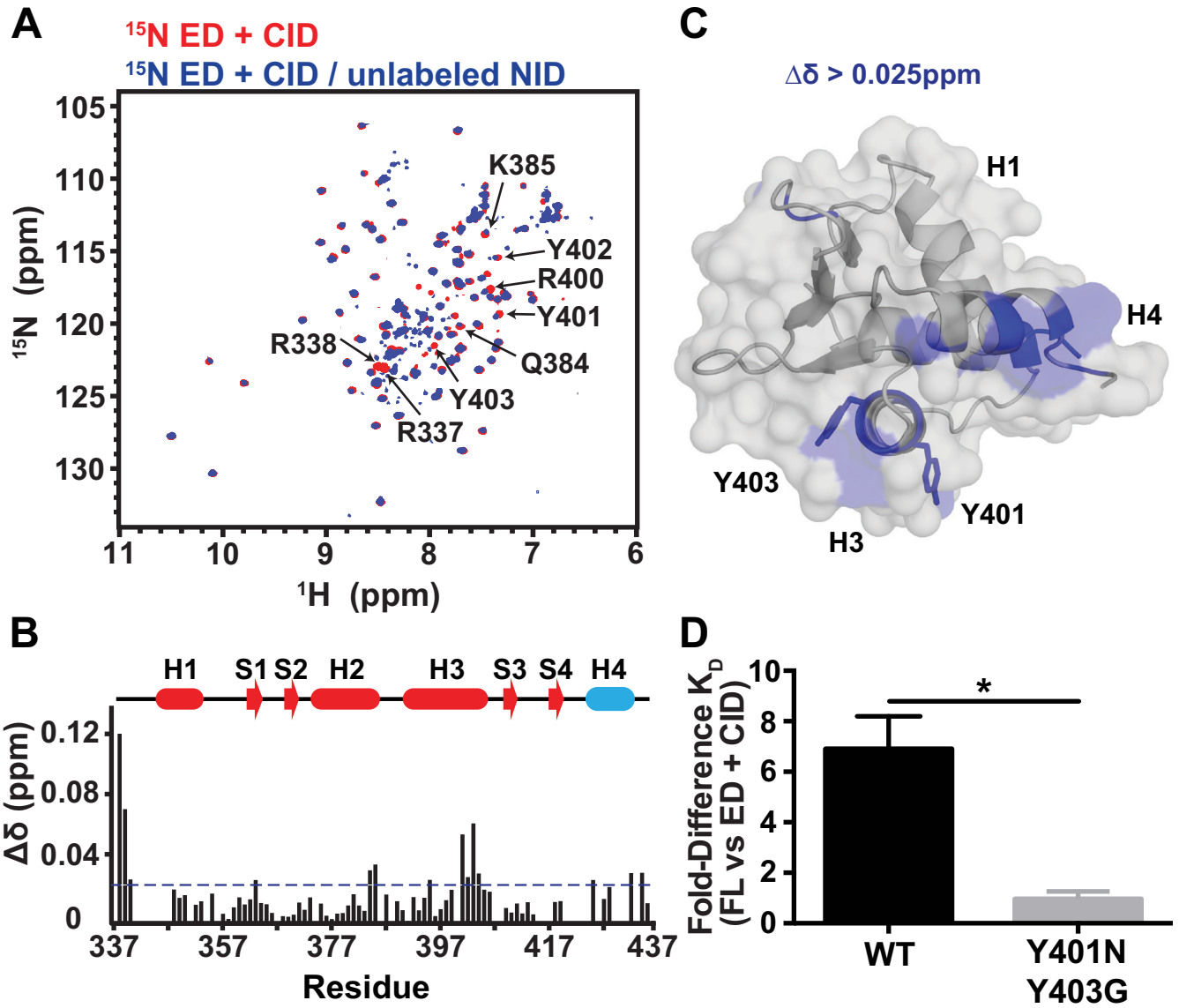


Figure 7. The NID interacts with the CID and the DNA-recognition helix H3. (**A**) Overlaid ^{15}N-HSQC spectra of ^{15}N-labeled ETV4 ETS domain and CID alone (337–436, red), and with the unlabeled NID (165–336, blue) joined via intein ligation. Selected peaks are labeled. (**B** and **C**) The amide chemical shift perturbations, ($\Delta\delta = [(\Delta\delta_H)^2 + (0.2\ \Delta\delta_N)^2\ ]^{\frac{1}{2}}$), resulting from the ligated NID are displayed in histogram format and mapped onto the structure of ETV4 ETS domain and CID (blue, $\Delta\delta > 0.025$ ppm; gray, $\Delta\delta < 0.025$ ppm, prolines, or residues with unassigned NMR signals). (**D**) Fold difference of K_D values between full length (FL: residues 1–484) and CID-inhibited (ED + CID: 337–436) ETV4 for wild-type proteins (WT) or proteins with both Y401N and Y403G mutations. Bars and error bars represent the mean and the standard error of the mean for the following number of replicates: FL Y401N/Y403G, 6; ED + CID Y401N/Y403G, 5. '*' indicates $P < 0.05$. See Supplementary Figure S10 for binding isotherms.

broad such that it is resilient to individual mutations, or the CID and the NID indirectly interact via the direct interactions between the NID and H3, and between H3 and the CID.

We took another approach to detect intramolecular interaction by following a biological regulatory activity. Widespread acetylation of lysine residues activates the DNA binding of ETV4, and two known sites of acetylation, Lys226 and Lys260, reside within the NID (49–51). Therefore, we tested whether acetylation of these residues is sufficient for activating ETV4 DNA binding. Acetylation of either Lys226 or Lys260, independently, resulted in a decrease of DNA binding autoinhibition by 2.8- or 1.6-fold, respectively (Figure 8A and B). First, we hypothesized that the positive charge of these lysine residues may be important for inhibition. However, mutation of Lys226 and Lys260 to glutamate or glutamine failed to recapitulate the activating nature by acetylation of these residues (Supplementary Figure S11). Next we tested whether hydrophobic forces provided intramolecular interactions between the NID and the ETS domain and CID, such that the added bulk of acetyllysine might disrupt such interactions formed by nearby aro-

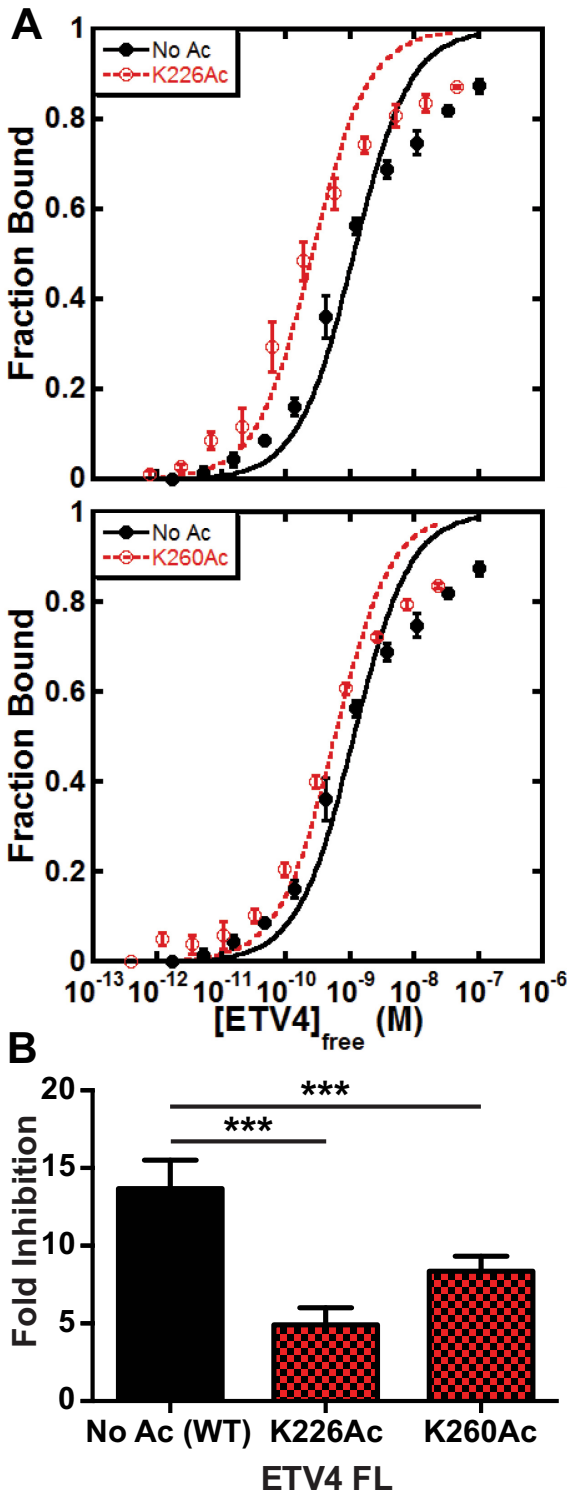


Figure 8. Acetylation of Lys226 or Lys260 relieves NID-dependent autoinhibition. (**A**) Binding isotherms for FL ETV4 in its unacetylated form (No Ac; black), and acetylated at Lys226 (top, red) or Lys260 (bottom, red). Data points and error bars correspond to the mean and standard error of the mean from four replicates. (**B**) Quantification of fold inhibition relative to uninhibited ETV4 (337–430) as in Figures 1D and 2A. The DNA binding of ETV4 Lys226Ac (K_D, $30 \pm 6 \times 10^{-11}$ M) and ETV4 Lys260Ac (K_D, $51 \pm 3 \times 10^{-11}$ M) was inhibited 5 ± 1-fold and 8 ± 1-fold, respectively, whereas, unmodified ETV4 was inhibited 14 ± 2-fold. '***' Indicates $P < 0.001$.

matic residues in the NID. Conserved aromatic residues that are located proximally to Lys226 in the NID were substituted with alanines, singly and in combination. However, this mutagenesis did not activate DNA binding (Supplementary Figure S11). Therefore, while the exact nature of the inhibiting residues within the NID remains unclear, several lines of evidence support the occurrence of intramolecular interactions that might include a steric mechanism of DNA-binding autoinhibition.

DISCUSSION

Mechanistic model of autoinhibition

Here, we demonstrated that members of the ETV1/4/5 subfamily of ETS factors have regions N- and C-terminal of their ETS domains that act together to impinge upon helix H3 and inhibit DNA binding (Figure 9). We propose that the CID allosterically shifts the conformational equilibrium of the DNA-recognition helix H3 toward a state less competent for binding. In contrast, the NID works, at least in part, in a steric manner to occlude the DNA binding interface, thus also requiring a conformational change for DNA binding.

The CID functions by influencing the position of helix H3, as supported by structural and mutational analysis. Amide HX experiments revealed that the CID helix H4 and the DNA-recognition helix H3 are both dynamic. These helices also exist in distinct conformations in crystallographic structures. Importantly, the uninhibited and the DNA-bound conformations of H3 match one another, but are distinct from the CID-inhibited conformation (45). We propose that the dynamic nature of helices H3 and H4, detected by HX measurements, reflect the sampling of these multiple conformations. Direct interactions between Ile407-Leu430 couple these two helices allowing the CID to 'push' H3 toward a state with lower affinity for DNA binding (Figure 10A).

The NID is predominantly intrinsically disordered and inhibits the ETS domain through interactions with helix H3, and possibly with the CID. These interactions are weak and transient as evidenced by the small NMR spectral perturbations accompanying ligation of the NID, the lack of detectable structural changes in the NID due to the presence of the ETS domain or CID, and the contribution of multiple regions of the NID to autoinhibition. However, multiple weak (or 'fuzzy') interactions (52) may lead to the overall inhibitory effect of the NID on the ETS domain and/or CID of ETV4. Tyr401 and Tyr403 in H3 are critical for NID-mediated autoinhibition and these residues directly contact DNA base pairs, suggesting that the NID sterically occludes part of the DNA binding interface. The NID also influences the CID, either through direct interaction that utilizes a broad interface or indirectly through the composite of the NID-H3 and H3-CID interactions. The tyrosine residues from H3 are not conserved in all ETS domains, and the sequence and positioning of CID helix H4 is unique to the ETV1/4/5 subfamily; therefore, the putative NID-interaction interface is specific to the DBD of this subfamily (Figure 10A). Based on these collective findings, we propose that ETV1, ETV4 and ETV5 are in a dynamic

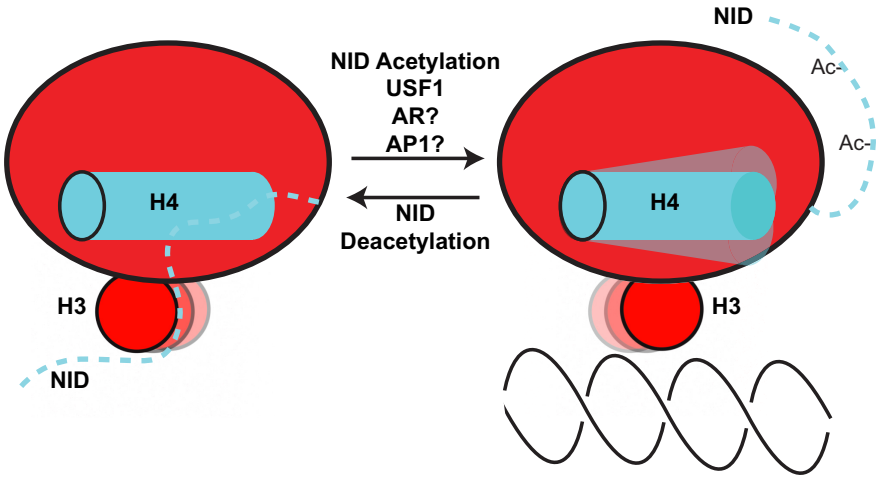


Figure 9. ETV1/4/5 subfamily factors are in equilibrium between forms that are more or less competent for binding to DNA. The N-terminal inhibitory domain (NID) and C-terminal inhibitory domain (H4) inhibit the ETS domain (red oval) of ETV1/4/5 subfamily factors. H4 'pushes' the DNA-recognition helix (H3) toward a position that is less competent for DNA binding. The NID makes direct contact with the DNA-recognition helix to sterically inhibit DNA binding, and may also reinforce the inhibitory position of H4. Acetylation of lysine residues in the NID partially relieves autoinhibition, likely by disrupting NID interactions with the ETS domain. USF1 relieves ETV4 autoinhibition (13), and we speculate that interactions with other factors, such as the DNA binding factors AR or AP1, may also regulate ETV4 DNA-binding activity.

equilibrium between conformations with different competencies for binding to DNA and that the NID and CID shift the equilibrium toward the less competent state (Figure 9).

The binding affinities of ETV1/4/5 fragments with either amino or carboxyl truncations suggests that the NID and CID work cooperatively, rather than additively, to inhibit DNA binding. Broadly speaking, this cooperation is supported by our structural and mutational data as the NID and the CID both impact the same location of the ETS domain, including the C-terminal portion of the DNA-recognition helix H3. In contrast to CID inhibition, there is insufficient understanding of the NID to ascertain the full basis of its inhibitory effects. Nevertheless, we speculate that in addition to a simple steric mechanism, as evidenced by direct perturbations of the DNA binding interface, the NID also reinforces the CID-driven conformation of helix H3. This is consistent with a cooperative mechanism of autoinhibition and the NID-induced perturbations of residues in both the ETS domain and the CID, detected by NMR spectroscopy.

Autoinhibition in ETS family of transcription factors

The characterization of autoinhibition in the ETV1/4/5 subfamily adds to the diversity of molecular mechanisms utilized in inhibiting DNA binding by ETS factors (Figure 10B). The DNA binding of ETS1 is allosterically inhibited by an α-helical module that flanks the ETS domain and by an IDR termed the serine-rich region (SRR). (7,8). Partial unfolding of the inhibitory module is linked to DNA binding (53,54). In contrast, a single C-terminal α-helix sterically inhibits the DNA-binding interface of ETV6 and unfolds to allow DNA binding (9–11). An α-helix and an IDR are reported to allosterically inhibit the DNA binding of ERG (43). Although these inhibitory elements show no sequence similarity (Supplementary Figure S1) and are structurally distinct between ETS factors (Figure 10), in all of these cases helix H4 interacts with a conserved hydrophobic surface on the ETS domain (Supplementary Figure S12). As this interaction is important for coupling the CID to the DNA-recognition helix H3 in ETV4, this conserved hydrophobic contact may reflect a common mechanism of inhibition among ETS factors. Beyond this conserved contact, the diversity of appended helices likely facilitates distinct intramolecular interactions with inhibitory IDRs and intermolecular interactions with unique protein partners.

Although the inhibitory domains of ETV1/4/5 are distinct from the previously characterized examples of ETS1, ETV6 and ERG, the cooperation of inhibitory elements is most reminiscent of ETS1 autoinhibition. Four α-helices flanking the ETS domain of ETS1 provide a small 2-fold level of inhibition (31), and this autoinhibition is reinforced to ∼20-fold by the SRR (8,12). As is the case for the proposed interaction between the NID and the ETS domain/CID of ETV4, the dynamic SRR also interacts transiently with both the flanking inhibitory α-helices of ETS1 and its ETS domain. Furthermore, tyrosine and phenylalanine residues, amino acids that are usually depleted within IDRs (55), are present in the SRR of ETS1 (56) and in the NID of ETV1, ETV4 and ETV5. However, in ETS1, these aromatic residues reside in a repeating Ser-(Tyr/Phe)-Asp pattern, and these repeats are key to the transient interactions that respond to signaling-induced phosphorylation and mediate inhibition. Such a repeat unit is not observed in the NID of ETV1/4/5, and mutation of aromatic residues in the NID of ETV4 did not influence autoinhibition. Therefore, although inhibitory IDRs are present in ETS1 and ETV1/4/5, these IDRs appear to modulate DNA binding via different intramolecular interactions with their corresponding ETS domains.

Divergent biological pathways and protein partnerships regulate the inhibitory elements of the ETS factors. Serine phosphorylation of ETS1, which targets the Ser-(Tyr/Phe)-

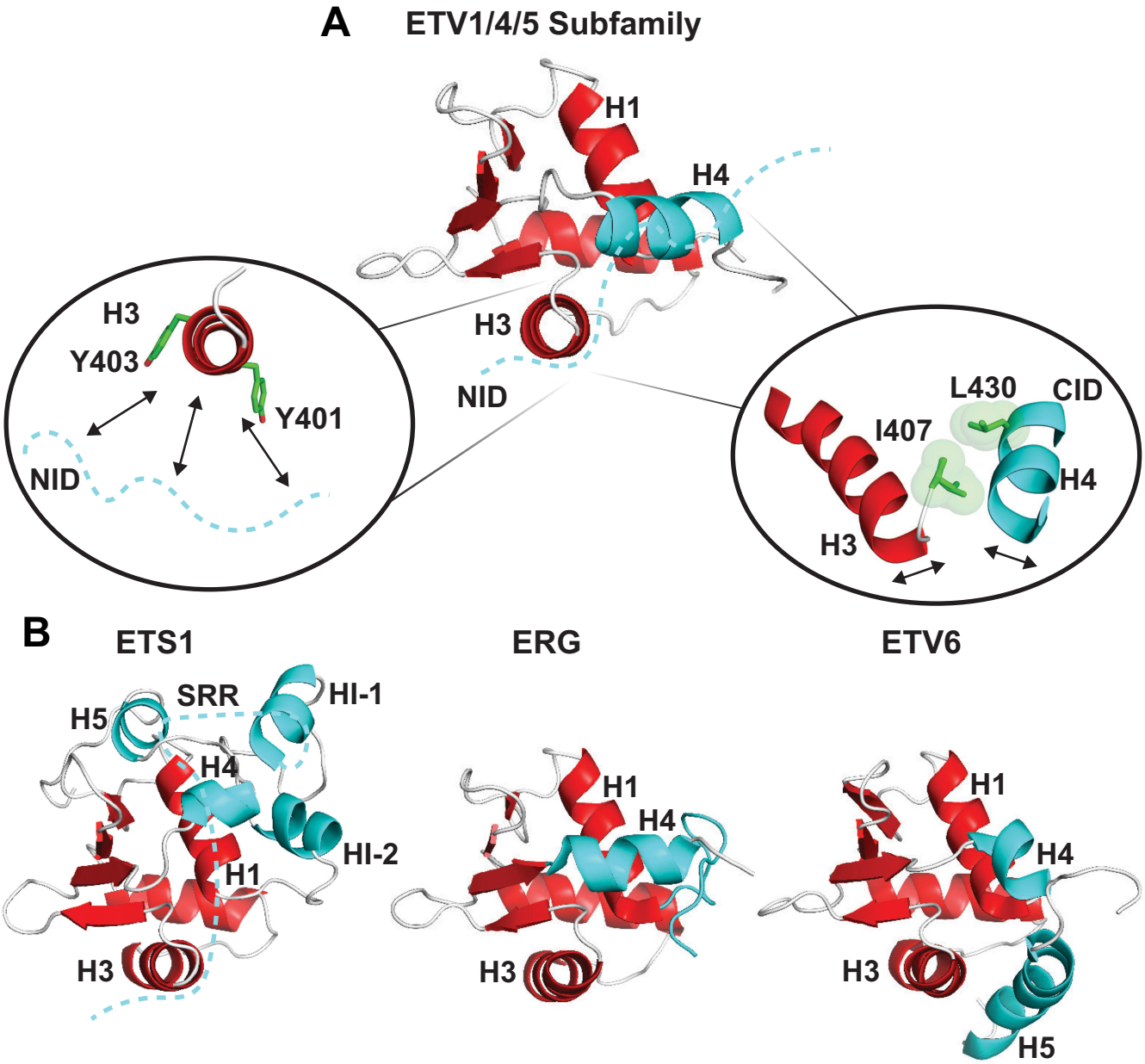


Figure 10. Autoinhibition in ETS family of transcription factors (ETS domain, red; inhibitory elements, cyan). (**A**) Structural and molecular elements of ETV1/4/5 subfamily autoinhibition. The CID is an α-helix, H4, that interacts with the ETS domain primarily through Leu430 (ETV4 numbering). In particular, Leu430 interacts with Ile407 to influence the positioning of H3 (right inset). The NID, cyan dotted line, is intrinsically disordered and interacts via multiple regions with H3 of the ETS domain, as well as the CID. Tyr401 and Tyr403 of H3 are required for NID-mediated autoinhibition (left inset). (**B**) Previous examples of structurally characterized autoinhibited ETS factors: ETS1 (7,8,12,14), ERG (43) and ETV6 (9–11). Dashed cyan lines refer to the disordered NID of ETV1/4/5 or the SRR of ETS1.

Asp repeat in the SRR, enhances the DNA-binding autoinhibition (8,56). In contrast, serine phosphorylation of ETV1 does not impact autoinhibition (57). Conversely, the relief of ETV4 autoinhibition by lysine acetylation, reported here, has not been observed for ETS1, ETV6 or ERG. Similarly, disparate protein partnerships regulate the DNA-binding autoinhibition of ETS1 and ETV1/4/5. For example, RUNX1 (15,58) and PAX5 (14) counter ETS1 autoinhibition, and USF1 relieves ETV4 autoinhibition (13). The RUNX1-ETS1 partnership results in ETS1-specific regulation at ETS-RUNX composite sites in T-cells (59), suggesting that the specific regulation of DNA-binding autoinhibition for an individual ETS factor can form the basis for that factor's unique biological function.

With a mechanistic foundation now in place for ETV1/4/5 autoinhibition, potential regulatory routes may be discovered, thus, providing insight into how these factors function in prostate cancer. We identified one possible route of ETV4 DNA binding activation through acetylation of lysines in the NID. Interestingly, the expression of p300, one of the acetyltransferases that modifies ETV1/4/5 factors (49,50), correlates with prostate cancer progression and high levels of p300 are prognostic of biochemical recurrence in prostate cancer patients (60,61). Additionally, protein-protein partnerships may regulate the DNA binding of ETV1/4/5 factors. Investigating the effect on ETV1/4/5 DNA-binding autoinhibition by other transcription factors that bind to proximal genomic sites and act in the prostate will be of particular interest. Besides USF1 (13), candidates include the AP1 factors (44) and the androgen receptor (22,62–64). Finally, inhibitory IDRs, and the signaling pathways that regulate them, are a potential therapeutic target as they differ between ETS factors and could provide factor-specific interventions. Despite the difficulty in rationally inhibiting IDRs (65), recent successes suggest that IDRs are a tractable small-molecule target (66–69).

Autoinhibition as a route to transcription factor specificity

Many transcription factors are encoded by gene families and share a conserved DBD (70). Disparate roles in development and disease indicate that individual transcription factors carry out specific functions and are not completely redundant with all other factors from the same family (6,71). ETS factors, as a prototype for investigating this conundrum, have provided insight into how such specificity could have evolved. Previous work had established that ETS factors have distinct inhibitory domains that are specific to an individual factor, or subfamily of factors (7–10,43). Cellular triggers can specifically regulate an individual ETS factor by integrating different signaling pathways or protein partnerships into these distinctive inhibitory features (14–16,58,59). We have extended this knowledge by describing an additional inhibitory module in the ETV1/4/5 subfamily with a distinct mechanism of inhibition and mode of cellular regulation. The divergent inhibitory domains of the four subclasses of ETS factors studied to date contact a conserved region on the ETS DBD. Thus, subfamily-specific α-helices that flank the ETS domain serve as 'adapters' that generate unique intra- and intermolecular interaction surfaces. Subfamily-specific IDRs interact with these surfaces to inhibit DNA binding via steric and/or allosteric mechanisms. These unstructured regions also provide diverse sites of post-translational modification that can inhibit or activate a factor in response to cellular regulation. Thus, the modest sequence variability among related DBDs could have been leveraged during evolution to enhance biological specificity by the coordinated divergence of structured inhibitory regions that flank the DBD and more distal inhibitory IDRs.

ACCESSION NUMBERS

Structural coordinates have been deposited in the RCSB Protein Data Bank under ID codes 5ILS (ETV1), 5ILU (ETV4) and 5ILV (ETV5).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Chris Hill and Dr Michael Kay for access to, and helpful discussions on, X-ray crystallography and circular dichroism, respectively. We acknowledge Dr Jason Chin, Dr Pierre Barraud and Dr Michael Sattler for providing plasmids for incorporating acetyllysine residues, intein ligation and sortase ligation, respectively. We thank Niraja Bhachech for assistance with protein purification and members of the Graves and McIntosh labs for helpful discussions.

FUNDING

National Institutes of Health [R01GM38663 to B.J.G.]; Canadian Cancer Society Research Institute (CCSRI) [2011–700772 to L.P.M.]; Canadian Institutes of Health Research (CIHR) [MOP-136834 to L.P.M.]; CIHR; Canada Foundation for Innovation; British Columbia Knowledge Development Fund; UBC Blusson Fund; Michael Smith Foundation for Health Research; Huntsman Cancer Institute/Huntsman Cancer Foundation and the Howard Hughes Medical Institute (to B.J.G.); National Cancer Institute, National Institutes of Health [P30CA042014 to Huntsman Cancer Institute]; National Institute of General Medical Sciences, National Institutes of Health [P50 GM082545 to Chris Hill and Wes Sundquist]; U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences [DE-AC02-76SF00515]; DOE Office of Biological and Environmental Research; National Institutes of Health, National Institute of General Medical Sciences [P41GM103393]. Funding for open access charge: National Institutes of Health [R01GM38663].
Conflict of interest statement. None declared.

REFERENCES

1. Pufall,M.A. and Graves,B.J. (2002) Autoinhibitory domains: modular effectors of cellular regulation. *Annu. Rev. Cell Dev. Biol.*, **18**, 421–462.
2. Kim,A.S., Kakalis,L.T., Abdul-Manan,N., Liu,G.A. and Rosen,M.K. (2000) Autoinhibition and activation mechanisms of the Wiskott-Aldrich syndrome protein. *Nature*, **404**, 151–158.
3. Morreale,A., Venkatesan,M., Mott,H.R., Owen,D., Nietlispach,D., Lowe,P.N. and Laue,E.D. (2000) Structure of Cdc42 bound to the GTPase binding domain of PAK. *Nat. Struct. Biol.*, **7**, 384–388.
4. Trudeau,T., Nassar,R., Cumberworth,A., Wong,E.T., Woollard,G. and Gsponer,J. (2013) Structure and intrinsic disorder in protein autoinhibition. *Structure*, **21**, 332–341.
5. Wright,P.E. and Dyson,H.J. (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.*, **16**, 18–29.
6. 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.
7. Lee,G.M., Donaldson,L.W., Pufall,M.A., Kang,H.S., Pot,I., Graves,B.J. and McIntosh,L.P. (2005) The structural and dynamic basis of Ets-1 DNA binding autoinhibition. *J. Biol. Chem.*, **280**, 7088–7099.
8. 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.

9. Green,S.M., Coyne,H.J. 3rd, McIntosh,L.P. and Graves,B.J. (2010) DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *J. Biol. Chem.*, **285**, 18496–18504.
10. Coyne,H.J. 3rd, De,S., Okon,M., Green,S.M., Bhachech,N., Graves,B.J. and McIntosh,L.P. (2012) Autoinhibition of ETV6 (TEL) DNA binding: appended helices sterically block the ETS domain. *J. Mol. Biol.*, **421**, 67–84.
11. De,S., Chan,A.C., Coyne,H.J. 3rd, Bhachech,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.
12. Lee,G.M., Pufall,M.A., Meeker,C.A., Kang,H.S., 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.
13. Greenall,A., Willingham,N., Cheung,E., Boam,D.S. and Sharrocks,A.D. (2001) DNA binding by the ETS-domain transcription factor PEA3 is regulated by intramolecular and intermolecular protein.protein interactions. *J. Biol. Chem.*, **276**, 16207–16215.
14. 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.
15. 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.
16. Shiina,M., Hamada,K., Inoue-Bungo,T., Shimamura,M., Uchiyama,A., Baba,S., Sato,K., Yamamoto,M. and Ogata,K. (2015) A novel allosteric mechanism on protein-DNA interactions underlying the phosphorylation-dependent regulation of Ets1 target gene expressions. *J. Mol. Biol.*, **427**, 1655–1669.
17. Tomlins,S.A., Rhodes,D.R., Perner,S., Dhanasekaran,S.M., Mehra,R., Sun,X.W., Varambally,S., Cao,X., Tchinda,J., Kuefer,R. *et al.* (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*, **310**, 644–648.
18. Tomlins,S.A., Mehra,R., Rhodes,D.R., Smith,L.R., Roulston,D., Helgeson,B.E., Cao,X., Wei,J.T., Rubin,M.A., Shah,R.B. *et al.* (2006) TMPRSS2:ETV4 gene fusions define a third molecular subtype of prostate cancer. *Cancer Res.*, **66**, 3396–3400.
19. Helgeson,B.E., Tomlins,S.A., Shah,N., Laxman,B., Cao,Q., Prensner,J.R., Cao,X., Singla,N., Montie,J.E., Varambally,S. *et al.* (2008) Characterization of TMPRSS2:ETV5 and SLC45A3:ETV5 gene fusions in prostate cancer. *Cancer Res.*, **68**, 73–80.
20. Tomlins,S.A., Laxman,B., Dhanasekaran,S.M., Helgeson,B.E., Cao,X., Morris,D.S., Menon,A., Jing,X., Cao,Q., Han,B. *et al.* (2007) Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature*, **448**, 595–599.
21. Aytes,A., Mitrofanova,A., Kinkade,C.W., Lefebvre,C., Lei,M., Phelan,V., LeKaye,H.C., Koutcher,J.A., Cardiff,R.D., Califano,A. *et al.* (2013) ETV4 promotes metastasis in response to activation of PI3-kinase and Ras signaling in a mouse model of advanced prostate cancer. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, E3506–E3515.
22. Baena,E., Shao,Z., Linn,D.E., Glass,K., Hamblen,M.J., Fujiwara,Y., Kim,J., Nguyen,M., Zhang,X., Godinho,F.J. *et al.* (2013) ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and patients. *Genes Dev.*, **27**, 683–698.
23. Laget,M.P., Defossez,P.A., Albagli,O., Baert,J.L., Dewitte,F., Stehelin,D. and de Launoit,Y. (1996) Two functionally distinct domains responsible for transactivation by the Ets family member ERM. *Oncogene*, **12**, 1325–1336.
24. Bojovic,B.B. and Hassell,J.A. (2001) The PEA3 Ets transcription factor comprises multiple domains that regulate transactivation and DNA binding. *J. Biol. Chem.*, **276**, 4509–4521.
25. Li,M.Z. and Elledge,S.J. (2012) SLIC: a method for sequence- and ligation-independent cloning. *Methods Mol. Biol.*, **852**, 51–59.
26. Neumann,H., Peak-Chew,S.Y. and Chin,J.W. (2008) Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat. Chem. Biol.*, **4**, 232–234.
27. Studier,F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.*, **41**, 207–234.
28. Gasteiger,E., Hoogland,C., Gattiker,A., Duvaud,S., Wilkins,M.R., Appel,R.D. and Bairoch,A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: Walker,JM (ed). *The Proteomics Protocols Handbook, Humana Press*. Humana Press, Totowa, pp. 571–607.
29. Barraud,P. and Allain,F.H. (2013) Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRMs influences the nucleic acid binding topology. *J. Biomol. NMR*, **55**, 119–138.
30. Freiburger,L., Sonntag,M., Hennig,J., Li,J., Zou,P. and Sattler,M. (2015) Efficient segmental isotope labeling of multi-domain proteins using Sortase A. *J. Biomol. NMR*, **63**, 1–8.
31. 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.
32. Otwinowski,Z. and Minor,W. (1997) Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
33. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
34. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
35. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
36. Chen,V.B., Arendall,W.B. 3rd, Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.
37. Greenfield,N.J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.*, **1**, 2876–2890.
38. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
39. Goddard,T.D. and Kneller,D.G. (1999) Sparky. 3rd edn. University of California, San Francisco.
40. Sattler,M., Schleucher,J. and Griesinger,C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Mag. Reson. Spect.*, **34**, 93–158.
41. Hwang,T.L., van Zijl,P.C. and Mori,S. (1998) Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR*, **11**, 221–226.
42. Bai,Y., Milne,J.S., Mayne,L. and Englander,S.W. (1993) Primary structure effects on peptide group hydrogen exchange. *Proteins*, **17**, 75–86.
43. Regan,M.C., Horanyi,P.S., Pryor,E.E. Jr, Sarver,J.L., Cafiso,D.S. and Bushweller,J.H. (2013) Structural and dynamic studies of the transcription factor ERG reveal DNA binding is allosterically autoinhibited. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 13374–13379.
44. Hollenhorst,P.C., Ferris,M.W., Hull,M.A., Chae,H., Kim,S. and Graves,B.J. (2011) Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. *Genes Dev.*, **25**, 2147–2157.
45. Cooper,C.D., Newman,J.A., Aitkenhead,H., Allerston,C.K. and Gileadi,O. (2015) Structures of the Ets protein DNA-binding domains of transcription factors Etv1, Etv4, Etv5, and Fev: determinants of DNA binding and redox regulation by disulfide bond formation. *J. Biol. Chem.*, **290**, 13692–13709.
46. Hafsa,N.E. and Wishart,D.S. (2014) CSI 2.0: a significantly improved version of the Chemical Shift Index. *J. Biomol. NMR*, **60**, 131–146.
47. Camilloni,C., De Simone,A., Vranken,W.F. and Vendruscolo,M. (2012) Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry*, **51**, 2224–2231.
48. Dyson,H.J. and Wright,P.E. (2002) Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.*, **12**, 54–60.
49. Goel,A. and Janknecht,R. (2003) Acetylation-mediated transcriptional activation of the ETS protein ER81 by p300, P/CAF, and HER2/Neu. *Mol. Cell. Biol.*, **23**, 6243–6254.
50. Guo,B., Panagiotaki,N., Warwood,S. and Sharrocks,A.D. (2011) Dynamic modification of the ETS transcription factor PEA3 by

sumoylation and p300-mediated acetylation. *Nucleic Acids Res.*, **39**, 6403–6413.
51. Kim,H.J., Kim,S.H., Yu,E.J., Seo,W.Y. and Kim,J.H. (2015) A positive role of DBC1 in PEA3-mediated progression of estrogen receptor-negative breast cancer. *Oncogene*, **34**, 4500–4508.
52. Fuxreiter,M. (2012) Fuzziness: linking regulation to protein dynamics. *Mol. Biosyst.*, **8**, 168–177.
53. Petersen,J.M., Skalicky,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 alpha helix. *Science*, **269**, 1866–1869.
54. 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.
55. Williams,R.M., Obradovi,Z., Mathura,V., Braun,W., Garner,E.C., Young,J., Takayama,S., Brown,C.J. and Dunker,A.K. (2001) The protein non-folding problem: amino acid determinants of intrinsic order and disorder. *Pac. Symp. Biocomput.*, **6**, 89–100.
56. Desjardins,G., Meeker,C.A., Bhachech,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.
57. Wu,J. and Janknecht,R. (2002) Regulation of the ETS transcription factor ER81 by the 90-kDa ribosomal S6 kinase 1 and protein kinase A. *J. Biol. Chem.*, **277**, 42669–42679.
58. 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 alpha2. *Mol. Cell. Biol.*, **20**, 81–90.
59. Hollenhorst,P.C., Shah,A.A., Hopkins,C. and Graves,B.J. (2007) Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. *Genes Dev.*, **21**, 1882–1894.
60. Debes,J.D., Sebo,T.J., Lohse,C.M., Murphy,L.M., Haugen,D.A. and Tindall,D.J. (2003) p300 in prostate cancer proliferation and progression. *Cancer Res.*, **63**, 7638–7640.
61. Isharwal,S., Miller,M.C., Marlow,C., Makarov,D.V., Partin,A.W. and Veltri,R.W. (2008) p300 (histone acetyltransferase) biomarker predicts prostate cancer biochemical recurrence and correlates with changes in epithelia nuclear size and shape. *Prostate*, **68**, 1097–1104.
62. Massie,C.E., Adryan,B., Barbosa-Morais,N.L., Lynch,A.G., Tran,M.G., Neal,D.E. and Mills,I.G. (2007) New androgen receptor genomic targets show an interaction with the ETS1 transcription factor. *EMBO Rep.*, **8**, 871–878.
63. Chen,Y., Chi,P., Rockowitz,S., Iaquinta,P.J., Shamu,T., Shukla,S., Gao,D., Sirota,I., Carver,B.S., Wongvipat,J. *et al.* (2013) ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. *Nat. Med.*, **19**, 1023–1029.
64. Shin,S., Kim,T.D., Jin,F., van Deursen,J.M., Dehm,S.M., Tindall,D.J., Grande,J.P., Munz,J.M., Vasmatzis,G. and Janknecht,R. (2009) Induction of prostatic intraepithelial neoplasia and modulation of androgen receptor by ETS variant 1/ETS-related protein 81. *Cancer Res.*, **69**, 8102–8110.
65. Zhang,Y., Cao,H. and Liu,Z. (2015) Binding cavities and druggability of intrinsically disordered proteins. *Protein Sci.*, **24**, 688–705.
66. Hammoudeh,D.I., Follis,A.V., Prochownik,E.V. and Metallo,S.J. (2009) Multiple independent binding sites for small-molecule inhibitors on the oncoprotein c-Myc. *J. Am. Chem. Soc.*, **131**, 7390–7401.
67. Krishnan,N., Koveal,D., Miller,D.H., Xue,B., Akshinthala,S.D., Kragelj,J., Jensen,M.R., Gauss,C.M., Page,R., Blackledge,M. *et al.* (2014) Targeting the disordered C terminus of PTP1B with an allosteric inhibitor. *Nat. Chem. Biol.*, **10**, 558–566.
68. Pop,M.S., Stransky,N., Garvie,C.W., Theurillat,J.P., Hartman,E.C., Lewis,T.A., Zhong,C., Culyba,E.K., Lin,F., Daniels,D.S. *et al.* (2014) A small molecule that binds and inhibits the ETV1 transcription factor oncoprotein. *Mol. Cancer Ther.*, **13**, 1492–1502.
69. Zhang,Z., Boskovic,Z., Hussain,M.M., Hu,W., Inouye,C., Kim,H.J., Abole,A.K., Doud,M.K., Lewis,T.A., Koehler,A.N. *et al.* (2015) Chemical perturbation of an intrinsically disordered region of TFIID distinguishes two modes of transcription initiation. *Elife*, **4**, e07777.
70. Vaquerizas,J.M., Kummerfeld,S.K., Teichmann,S.A. and Luscombe,N.M. (2009) A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.*, **10**, 252–263.
71. Rezsohazy,R., Saurin,A.J., Maurel-Zaffran,C. and Graba,Y. (2015) Cellular and molecular insights into Hox protein action. *Development*, **142**, 1212–1227.

Structured and disordered regions cooperatively mediate DNA-binding autoinhibition of ETS factors ETV1, ETV4, and ETV5

Simon L. Currie[1,3] , Desmond K. Lau[4], Jedediah J. Doane[1,3], Frank G. Whitby[2], Mark Okon[4], Lawrence P. McIntosh[4], and Barbara J. Graves[1,3,5,*]

[1] Department of Oncological Sciences, [2] Department of Biochemistry, University of Utah School of Medicine, [3] Huntsman Cancer Institute, University of Utah, Salt Lake City UT, 84112-5550, USA,

[4] Department of Biochemistry and Molecular Biology, Department of Chemistry, and Michael Smith Laboratories, University of British Columbia, Vancouver BC, V6T 1Z3, Canada.[2]

[5] Howard Hughes Medical Institute, Chevy Chase MD, 20815-6789

* To whom correspondence should be addressed. Tel: 1-301-215-8718; Fax: 1-301-215-8828; Email: Barbara.Graves@hci.utah.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

SUPPLEMENTARY DATA

Supplementary Table S1.
X-ray crystallography data collection and refinement statistics

	ETV1 (332-435)	ETV4 (337-441)	ETV5 (364-457)
Data Collection			
Processing software	HKL2000	HKL2000	HKL2000
Beamline	SSRL 7-1	SSRL 7-1	SSRL 7-1
Wavelength	1.0000	1.0000	1.1271
Detector type	Q315 CCD	Q315 CCD	Q315 CCD
Collection date	2/7/13	2/7/13	1/12/13
Space group	$P3_121$	$P3_121$	$C222_1$
Unit cell	(50.2, 50.2, 69.3)	(50.9, 50.9, 68.6)	(57.5, 65.7, 53.0)
Resolution (Å)	55.00 - 1.40	45.00 - 1.10	30.00 - 1.80
Resolution (Å)	1.45 - 1.40	1.13 - 1.10	1.86 - 1.80
# Reflections measured	705,596	1,577,832	50,220
# Unique reflections	20,493	42,215	9,566
Redundancy	34.4	37.4	5.2
Completeness (%)	100.0 (100.0)[a]	100.0 (100.0)[a]	99.2 (97.3)[a]
<I/σI>	16 (1.9)[a]	5 (0.9)[a]	9 (1.0)[a]
Mosaicity (°)	0.4	0.2	1.3
R(pim)	0.018 (0.243)[a]	0.020 (0.676)[a]	0.039 (0.363)[a]
Refinement			
Refinement software	PHENIX.REFINE	PHENIX.REFINE	PHENIX.REFINE
Resolution (Å)	30.0 - 1.40	45.00 - 1.10	30.0 - 1.80
Resolution (Å)	1.47 - 1.40	1.13 - 1.10	2.05 - 1.80
# Reflections used for refinement	20,457	42,112	8163
# Reflections in R_{free} set	967	1,988	410
R_{cryst}	0.157 (0.217)[a]	0.181 (0.361)[a]	0.186 (0.247)[a]
R_{free}	0.178 (0.237)[a]	0.201 (0.388)[a]	0.234 (0.285)[a]
RMSD: bonds (Å) / angles (°)	0.006 / 1.175	0.005 / 1.047	.008 / 1.456
<B> ($Å^2$): All protein atoms / # atoms	16.1 / 890	16.5 / 1013	29.7 / 851
<B> ($Å^2$): water molecules / # water	32.8 / 114	28.9 / 125	37.1 / 81
Ramachandran favored (%)	87.5	91.8	87.7
Ramachandran additionally allowed (%)	12.5	8.2	12.3
Protein Data Bank ID	5ILS	5ILU	5ILV

[a]Values in parentheses are for highest-resolution shell.
One crystal was used to measure the data for each structure.

Supplementary Table S2.
Equilibrium dissociation constants (K_D) and fold-inhibition values for ETS factors

ETS Factor	Fragment	K_D (x10^{-11} M)[a]	Fold Inhibition[a,b]	*p*[c]	n
ERG	DBD (307-400)	40 ± 10	1.0 ± 0.5	-	3
	FL (1-479)	94 ± 9	2.3 ± 0.9	0.05	3
FLI1	DBD (277-370)	26 ± 8	1.0 ± 0.4	-	7
	FL (1-452)	70 ± 20	3 ± 1	0.1	3
ETV1	DBD (332-425)	5.4 ± 1.0	1.0 ± 0.3		6
	FL (1-479)	110 ± 20	21 ± 6	0.0006	10
ETV4	DBD (337-430)	6.1 ± 0.6	1.0 ± 0.1	-	25
	FL (1-484)	83 ± 8	14 ± 2	3 x 10^{-7}	35
ETV5	DBD (364-457)	3.6 ± 0.4	1.0 ± 0.2	-	4
	FL (1-510)	140 ± 30	39 ± 9	0.003	8
ETS1[d]	DBD	1.1 ± 0.1	1.0 ± 0.1	-	3
	FL	32 ± 4	29 ± 4	0.002	3
ETV6[e]	DBD	280 ± 40	1.0 ± 0.2	-	4
	FL	2,800 ± 400	10 ± 2	0.004	4

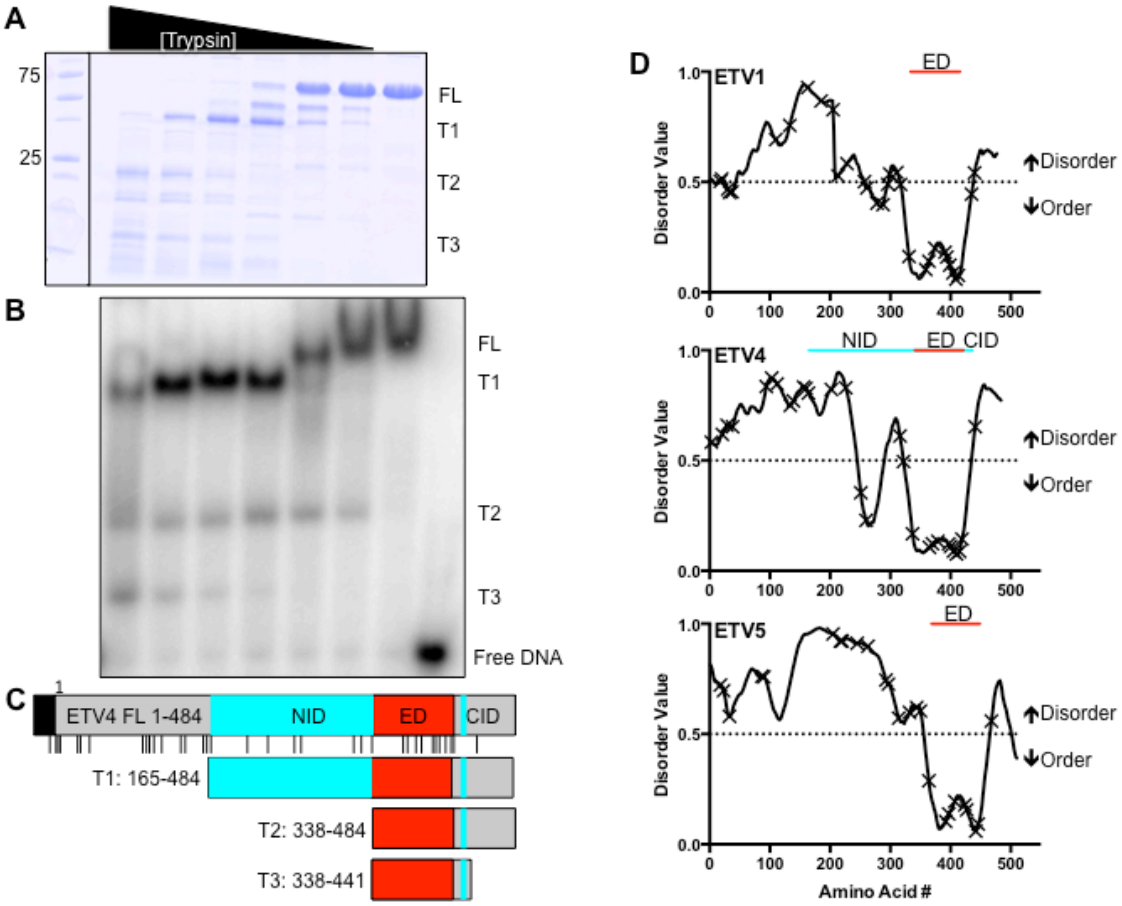
[a] Mean and standard error of the mean are given for K_D and fold-inhibition values.
[b] The DBD is set as uninhibited and used as a reference for calculating fold inhibition as K_D (FL or DBD) / K_D (DBD).
[c] The *p*-values were calculated using a two-tailed heteroscedastic t-test and compare the DBD and FL fragments for each ETS factor.
[d] Data included for comparison from supplementary reference (1).
[e] Data included for comparison from supplementary reference (2).

Supplementary Table S3.
Equilibrium dissociation constants (K_D) and fold inhibition values for ETV4 fragments

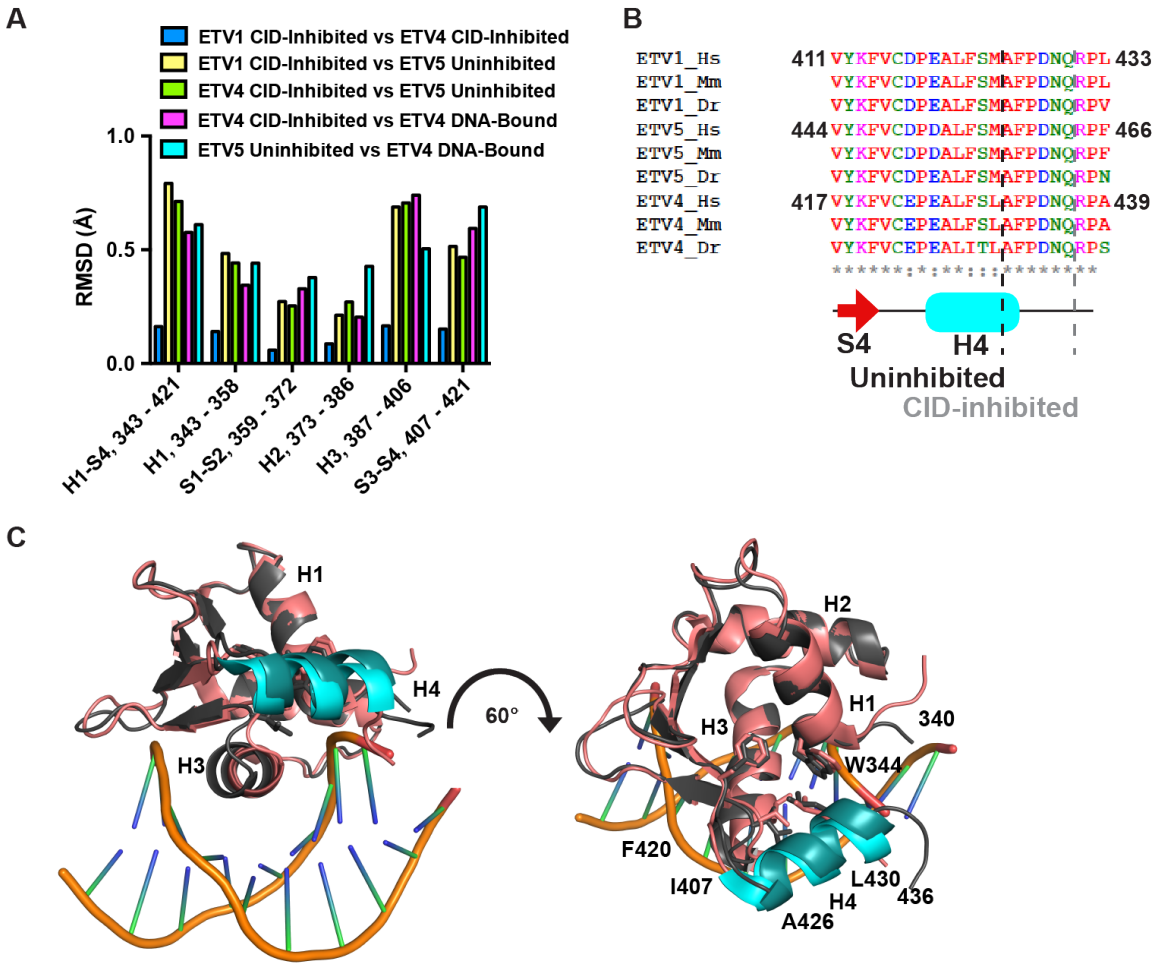
ETV4 Fragment	K_D (x10^{-11} M)[a]	Fold Inhibition[a,b]	*p*[c]	n
337-430 (DBD)	6.1 ± 0.6	1.0 ± 0.1	-	25
337-436	12 ± 2	2.0 ± 0.4	0.009	23
337-484	11 ± 2	1.8 ± 0.4	0.04	4
165-430	12 ± 1	2.1 ± 0.2	0.03	3
165-436	60 ± 10	10 ± 3	3 x 10^{-7}	11
165-484 (T1)	66 ± 9	11 ± 2	3 x 10^{-7}	18
1-484 (FL)	83 ± 8	14 ± 2	4 x 10^{-7}	35

[a] Mean and standard error of the mean are given for K_D and fold-inhibition values.
[b] ETV4 (337-430), the uninhibited DBD fragment, was used as a reference for calculating fold inhibition as K_D (fragment or full length) / K_D (ETV4 337-430).
[c] The *p*-values were calculated with ETV4 (337-430) as the reference.

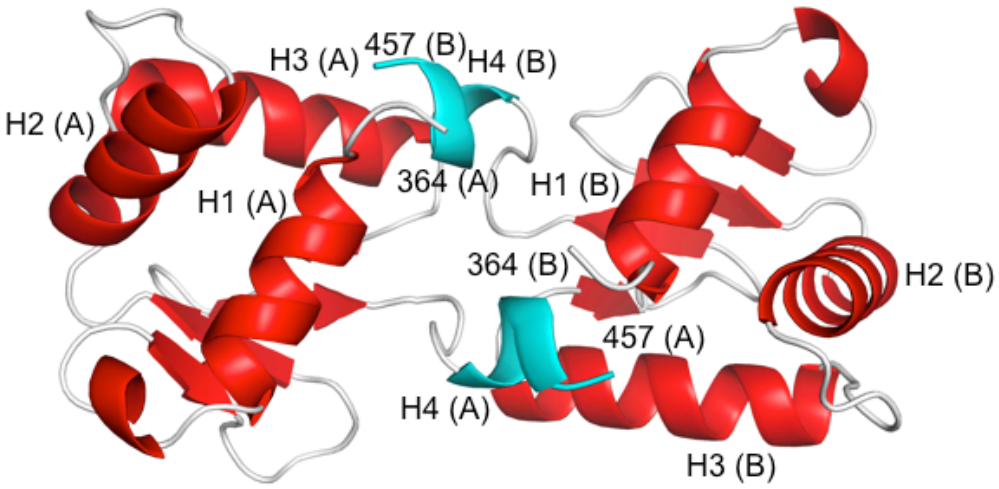
Supplementary Figure S1. Sequence alignment of ETS factors tested for autoinhibition. The full-length sequences for ETS factors tested for autoinhibition were aligned using Clustal Omega (3). Sequences for ETV7 and ETS2 were included as these factors belong to the same subfamilies as ETV6 and ETS1, respectively. The ETS domain (ETV4 residues 339-420) is shaded red and flanking α-helices and known inhibitory regions are shaded cyan and labeled per previous nomenclature (1,4-6). Residues discussed in this study are indicated by an arrow; ETV4 Y220, F225, K226, Y229, L233, Y234, W344, Y401, Y403, I407, F414, A426, and L430. These factors are highly conserved within the ETS domain, and are highly divergent outside of the ETS domain. Additionally, known inhibitory regions from ETV6 (H4 and H5) and ETS1 (SRR, HI-1, HI-2, H4, and H5) are not conserved in the ETV1/4/5 or ERG subfamilies.



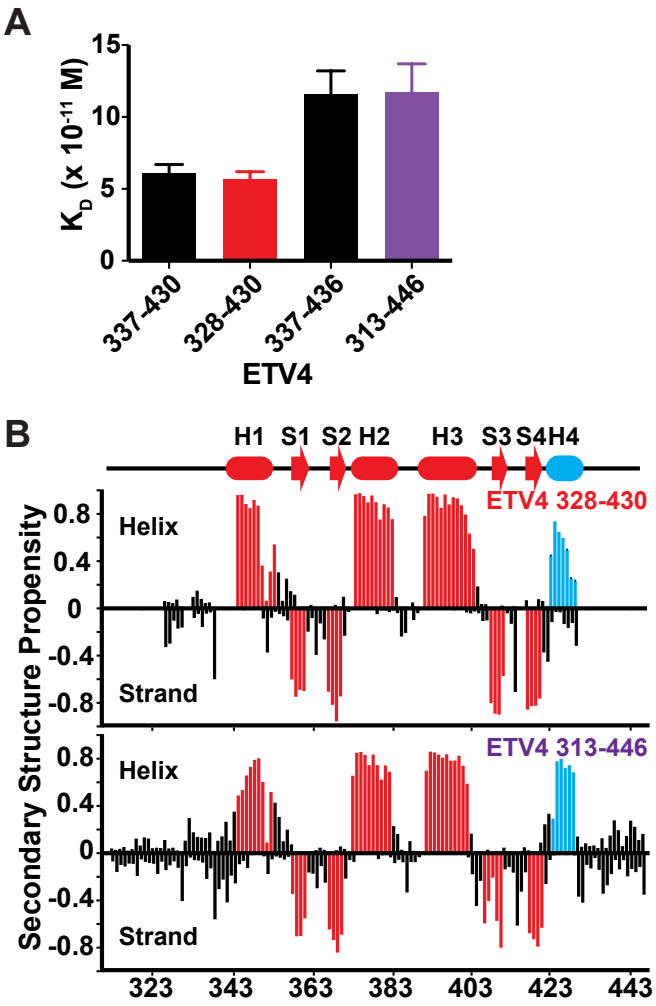
Supplementary Figure S2. ETV4 (165-484) is a trypsin-resistant fragment. (**A**) SDS-PAGE gel of partial trypsin proteolysis of ETV4. The leftmost lane contains protein molecular weight standards, and next seven lanes show products from two minute digestion with 450, 150, 45, 15, 4.5, 1.5, and 0 ng of trypsin. A representative example of three independent experiments is displayed. (**B**) Electrophoretic mobility shift assay with tryptic fragments from (A). The far-right lane is a DNA-only control. (**C**) Schematic of ETV4 full length (FL) and tryptic fragments retaining the ETS domain as identified by electrospray ionization mass spectrometry (ESI-MS). The predominant DNA-binding tryptic fragments are named T1, T2, and T3. The black bar refers to an N-terminal His_6-tag encoded by the pET28 vector and the vertical lines mark potential trypsin digestion sites as predicted by ExPASY Peptide Cutter (7). The ETS domain (ED) is noted in red, and N-terminal inhibitory domain (NID) and C-terminal inhibitory domain (CID), as identified for ETV4 (Figure 2), are noted in cyan. (**D**) Predicted disorder values are plotted over the full length of ETV1 (top), ETV4 (middle), and ETV5 (bottom). These values, calculated using Predictor of Naturally Disordered Regions (PONDR) VL3 (8), range from 0 (likely ordered) to 1 (likely disordered). Potential trypsin digestion sites are denoted by "X". Red lines refer to residues that span the ETS domain (ED), cyan lines in ETV4 refer to the NID and CID as identified for ETV4 (Figure 2).



Supplementary Figure S3. Structural comparison of CID-inhibited ETV1 and ETV4 with uninhibited ETV5 and DNA-bound ETV4. (**A**) Root mean square deviations (RMSDs) were calculated for backbone atoms to compare the crystal structures of uninhibited ETV5 (amino acids 364-457), with CID-inhibited ETV1 (331-435), and ETV4 (337-441), and DNA-bound ETV4 (337-441) (4UUV.pdb) (9). Secondary structural elements are defined as in Figure 3 and the numbering on the x-axis refers to ETV4. For subsections of the entire structure (e.g., H1, 343-358), the different structures were realigned based on that particular subsection and RMSD values correspond to backbone atoms within that subsection. The CID-inhibited ETV1 and ETV4 structures are very similar and have low RMSD values. The ETS domain overall (H1-S4), as well as most subsections (H1, S1-S2, H2, and S3-S4), have similar RMSD values for the remaining comparisons. In contrast, the RMSD value for H3 is lower for the uninhibited ETV5 versus DNA-bound ETV4 comparison than for the CID-inhibited ETV1/ETV4 versus uninhibited ETV5 or the CID-inhibited ETV4 versus DNA-bound ETV4 comparisons. This indicates that H3 is more similar in the uninhibited and DNA-bound states than in the CID-inhibited state. (**B**) Sequence alignment of ETV1/4/5 helix H4 from *H. sapiens* (Hs)*, M. musculus* (Mm)*, and D. rerio* (Dr) colored according to Clustal Omega (3). The red arrow and cyan cylinder indicate β-strand S4 of the ETS domain and α-helix H4, respectively. The vertical dashed black and gray lines identify truncation endpoints that cause activation or retain CID inhibition, respectively. (**C**) CID-inhibited ETV4 in its free (this study) and DNA-bound forms (4UUV.pdb) (9) were aligned based on the entire protein sequence. ETS domain and inhibitory residues are colored gray and dark teal, respectively, for the free ETV4 and pink and cyan, respectively, for the DNA-bound ETV4. Selected side chains are displayed in stick format as in Figure 3. Comparison with the free form demonstrates that there are subtle shifts of backbone atoms in the C-terminus of α-helix H3, as well as H4.



Supplementary Figure S4. Crystal packing of uninhibited ETV5 (364-457). The labels (A) and (B) distinguish the two molecules of uninhibited ETV5. The contacts between (A) and (B) may affect the position of truncated α-helix H4 (cyan) as compared to the position in solution or in the intact H4 in inhibited ETV5.

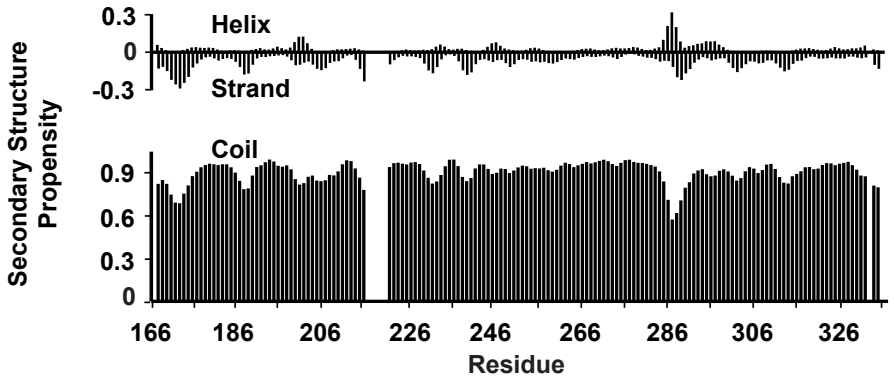


Supplementary Figure S5. ETV4 fragments used for NMR spectroscopic studies have the same affinities for DNA and secondary structures as similar sized fragments used for X-ray crystallography. (**A**) K_D values for the uninhibited (328-430, n=4, red) and inhibited (313-446, n=4, purple) ETV4 fragments used for NMR spectroscopy compared to those used for X-ray crystallography, black [n=25 and n=23 for ETV4 (337-430) and ETV4 (337-436), respectively]. (**B**) Secondary structure propensities for the two NMR-characterized ETV4 fragments calculated from their $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, ^{13}CO chemical shifts using the algorithm MICS (10). Helix, strand (shown as negative values), and coil (not shown) propensities sum to 1. Colored histogram bars identify amides in helices or strands of the ETS domain, red, and CID, cyan, as observed in the X-ray crystal structure of inhibited ETV4 (337-441) (Figure 5B). Although truncated, residues corresponding to H4 still adopt a folded α-helical conformation when ETV4 (328-430) is in the solution conditions used for NMR spectroscopic studies. However, the chemical shift-derived helical propensities of residues towards the C-terminus of the truncated H4 are reduced relative to the full helix in ETV4 (337-441). Amide ^{15}N relaxation measurements (not shown) also indicate that the C-terminal residues of ETV4 (328-430) are more mobile than those in the N-terminal portion of the truncated H4.

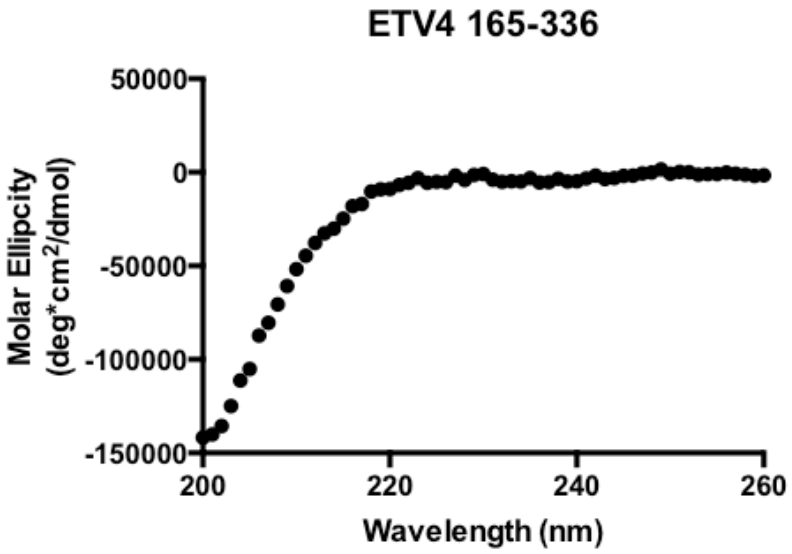
A

```
166  SGTSQPHPGHGYLGEHSSVFQQPLDICHSFTSQGGGREPLPAPYQHQLSE  215
     CCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHCCCCCCCCCCCCCCCCCC

216  PCPPYPQQSFKQEYHDPLYEQAGQPAVDQGGVNGHRYPGAGVVIKQEQTD  265
     CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

266  FAYDSDVTGCASMYLHTEGFSGPSPGDGAMGYGYEKPLRPFPDDVCVVPE  315
     CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC

316  KFEGDIKQEGVGAFREGPPYQ  336
     CCCCCCCCCCCCCCCCCCCCC
```

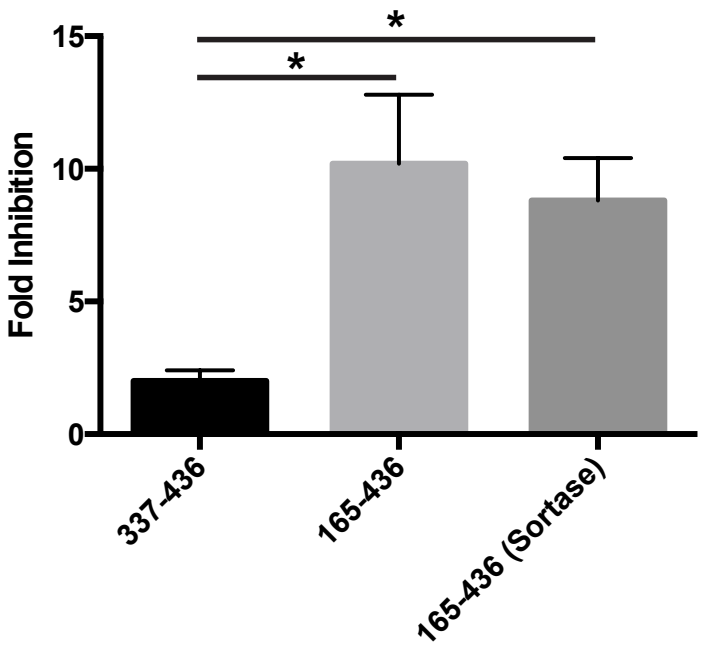
B



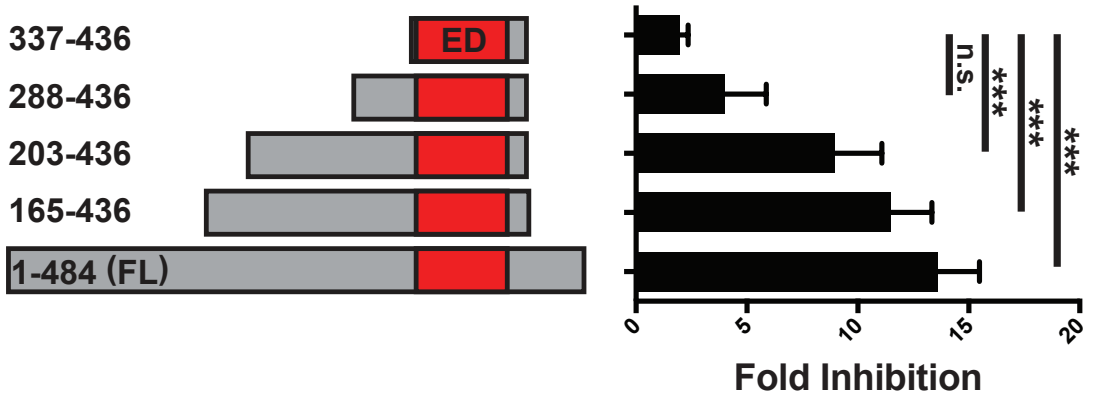
Supplementary Figure S6. The NID is intrinsically disordered. (**A**) Secondary structure propensities calculated from the assigned mainchain chemical shifts ($^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, ^{13}CO) of the isolated ETV4 NID (165-336) with CSI 2.0 (C, coil; H, helix) (11). (**B**) Normalized secondary structure propensities for α-helical (top, positive values), β-strand (top, negative values) and random coil or polyproline-II conformations (bottom), calculated from main chain chemical shifts using δ2D (12). Although differing in scoring criteria and output format, both algorithms reveal that the NID is predominantly disordered.



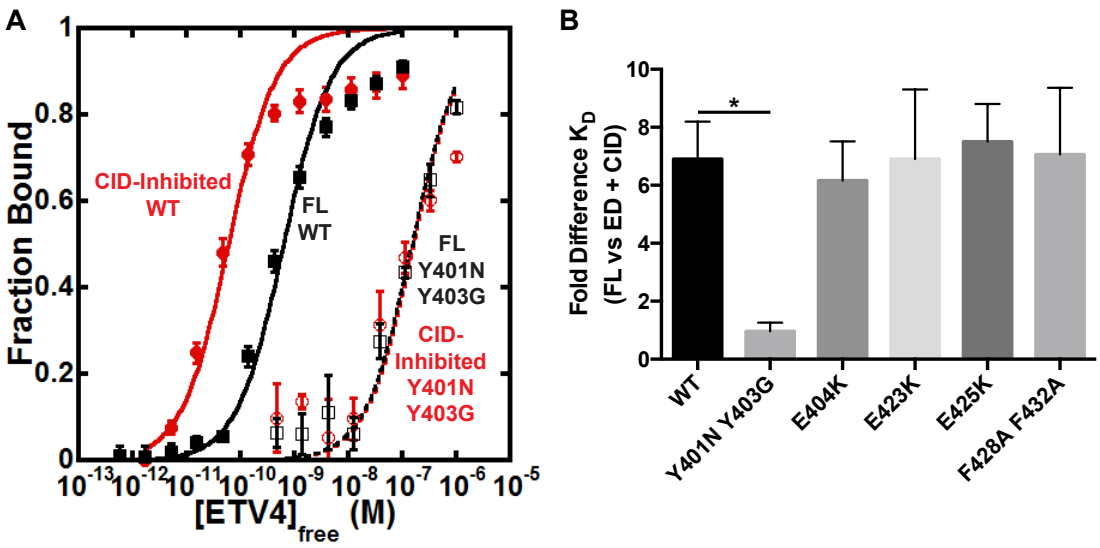
Supplementary Figure S7. The NID is intrinsically disordered. The circular dichroism spectrum of the ETV4 NID (165-336), at 4 °C and pH 7.9, is indicative of a random coil conformation. Three scans were collected in series and averaged after visually verifying their consistency.



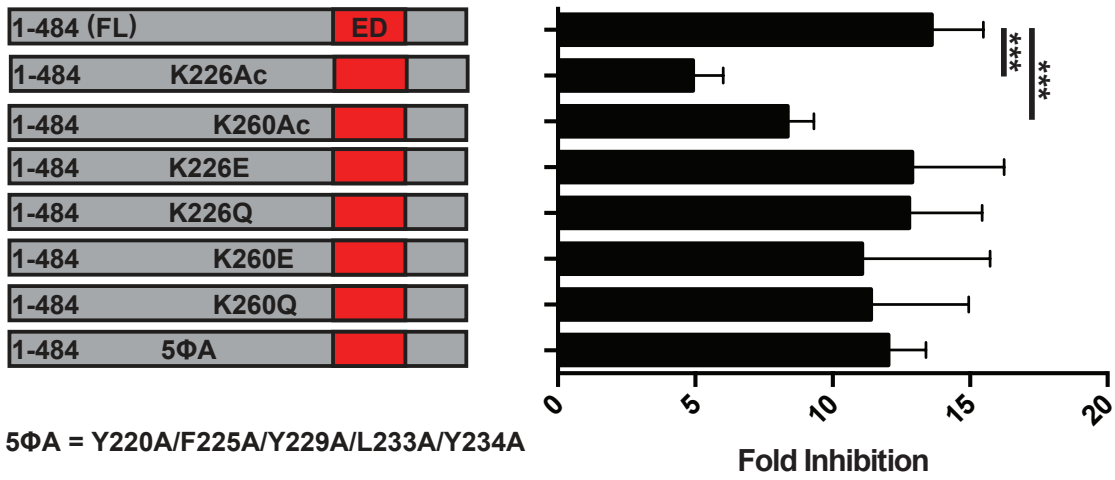
Supplementary Figure S8. Sortase-linked ETV4 (165-436) retains autoinhibition. Fold inhibition for ETV4 (337-436; ED + CID) versus (165-436; NID + ED + CID) expressed as a single protein or generated by sortase ligation of independently expressed (165-336) and (337-436) fragments. "*" indicates $p < 0.05$. These data indicate that the process of sortase ligation does not disrupt the autoinhibition of ETV4 (165-436).



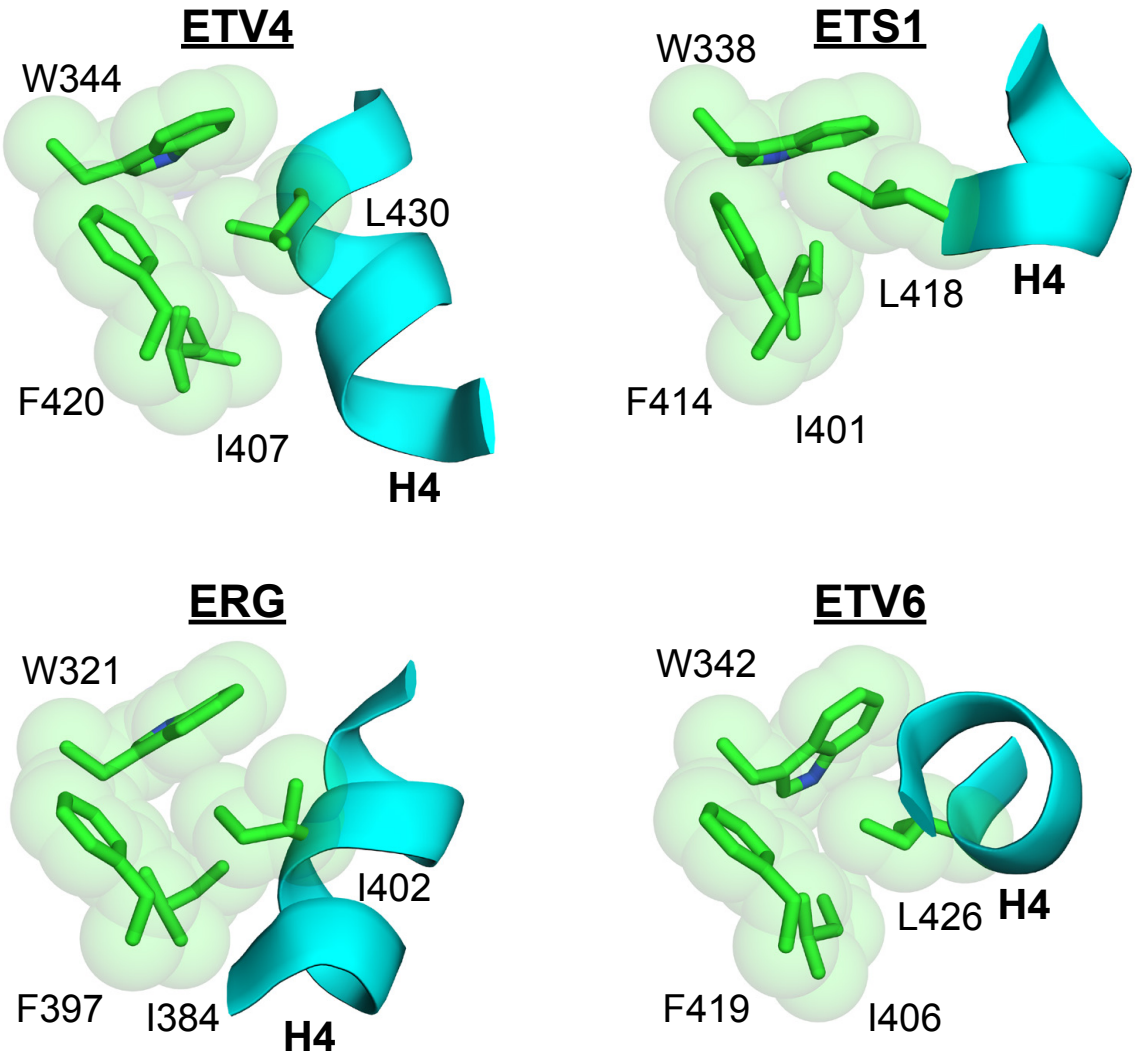
Supplementary Figure S9. Multiple regions within the NID contribute to the autoinhibition of ETV4. Fold-inhibition values for ETV4 fragments with various N-terminal truncations of the NID. Fold inhibition was calculated by comparing proteins to uninhibited ETV4 (337-430), as in Figures 1A and 2D. Bars and error bars refer to the mean and the standard error of the mean for the following number of replicates: 337-436, n = 23; 288-436, 8; 203-436, 8; 165-436, 11; 1-484, 35. "***" indicates $p < 0.001$ and n.s. indicates $p > 0.05$.



Supplementary Figure S10. Tyr401 and Tyr403 in H3 are required for NID-mediated inhibition. (**A**) Binding isotherms of CID-Inhibited ETV4 (red: 337-436) and full-length ETV4 (black: 1-484), solid lines and data points, and the same fragments with Y401N and Y403 point mutations, dashed lines and open data points. Data points and error bars represent the mean and the standard error of the mean for the following number of replicates: CID-inhibited, 7; FL, 14; CID-inhibited Y401N/Y403G, 3; FL Y401N/Y403G, 4. Mutating Tyr401 and Tyr403 weakens the affinity for DNA (compare CID-inhibited and CID-inhibited Y401N/Y403G). Importantly, these mutations also relieve autoinhibition from the NID (compare CID-inhibited Y401N/Y403G and FL Y401N/Y403G). (**B**) Fold difference of K_D values comparing full-length ETV4 (FL) and CID-inhibited ETV4 (337-436, ED + CID) for wild type (WT) proteins and with the indicated point mutants. Data points and error bars represent the mean and the standard error of the mean for the following number of replicates: WT, 23; Y401N Y403G, 5; E404K, 4; E423K, 4; E425K, 4; F428A F432A, 6. "*" indicates $p < 0.05$. These data demonstrate that Tyr401 and Tyr403 are required for NID-mediated autoinhibition and suggest that these residues are critical interaction sites for the NID. Although not altering autoinhibition, the E404K mutation strengthens DNA binding equally for both FL and CID-inhibited ETV4, whereas the E423K, E425K, and F428A/F432A substitutions have no effect on DNA binding (data not shown).



Supplementary Figure S11. Acetylation at Lys226 and Lys260 activates the DNA binding of ETV4. Fold-inhibition values for ETV4 with the indicated acetylated lysine residues and point mutations. Fold inhibition was calculated by comparing proteins to uninhibited ETV4 (337-430), as in Figures 1A and 2D. Bars and error bars refer to the mean and the standard error of the mean for the following number of replicates: 1-484, 35; 1-484 K226Ac, 4; 1-484 K260Ac, 4; K226E, 6; K226Q, 9; K260E, 3; K260Q, 7; Y220A F225A Y229A L233A Y234A, 3. "***" indicates $p < 0.001$.



Supplementary Figure S12. H4 is distinct in different ETS factors, but makes similar hydrophobic contacts with the ETS domain. ETV4 residues Trp344, Ile407, and Phe420 from the ETS domain and Leu430 from α-helix H4 are shown in van der Waals sphere format to illustrate the hydrophobic contacts between H4 and the ETS domain. ETS1, ERG, and ETV6 are formatted in the same way and shown at the same angle. In the ETV1/4/5 subfamily of factors we propose that the Ile407-Leu430 interaction inhibits DNA-binding by the "pushing" the DNA-recognition helix H3 into a conformation that is less competent for binding to DNA (Figure 4). The distinct, and subfamily-specific, versions of helix H4 in other ETS factors make similar hydrophobic contacts with this conserved surface on the ETS domain. Therefore, H3-H4 coupling may be a conserved mechanism of autoinhibition in ETS factors.

SUPPLEMENTARY REFERENCES

1. 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.
2. Green, S.M., Coyne, H.J., 3rd, McIntosh, L.P. and Graves, B.J. (2010) DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. *J Biol Chem*, **285**, 18496-18504.
3. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, **7**, 539.
4. 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.
5. Coyne, H.J., 3rd, De, S., Okon, M., Green, S.M., Bhachech, N., Graves, B.J. and McIntosh, L.P. (2012) Autoinhibition of ETV6 (TEL) DNA binding: appended helices sterically block the ETS domain. *J Mol Biol*, **421**, 67-84.
6. Regan, M.C., Horanyi, P.S., Pryor, E.E., Jr., Sarver, J.L., Cafiso, D.S. and Bushweller, J.H. (2013) Structural and dynamic studies of the transcription factor ERG reveal DNA binding is allosterically autoinhibited. *Proc Natl Acad Sci U S A*, **110**, 13374-13379.
7. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. (2005) In Walker, J. M. (ed.), *The Proteomics Protocols Handbook, Humana Press*. Humana Press, pp. 571-607.
8. Radivojac, P., Obradovic, Z., Brown, C.J. and Dunker, A.K. (2003) Prediction of boundaries between intrinsically ordered and disordered protein regions. *Pac Symp Biocomput*, 216-227.
9. Cooper, C.D., Newman, J.A., Aitkenhead, H., Allerston, C.K. and Gileadi, O. (2015) Structures of the Ets protein DNA-binding domains of transcription factors Etv1, Etv4, Etv5, and Fev: determinants of DNA binding and redox regulation by disulfide bond formation. *J Biol Chem*, **290**, 13692-13709.
10. Shen, Y. and Bax, A. (2012) Identification of helix capping and β-turn motifs from NMR chemical shifts. *J Biomol NMR*, **52**, 211-232.
11. Hafsa, N.E. and Wishart, D.S. (2014) CSI 2.0: a significantly improved version of the chemical shift index. *J Biomol NMR*, **60**, 131-146.
12. Camilloni, C., De Simone, A., Vranken, W.F. and Vendruscolo, M. (2012) Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry*, **51**, 2224-2231.