**SUPPLEMENTARY MATERIAL**

# Auto-inhibition of ETV6 (TEL) DNA-binding: appended helices sterically block the ETS domain

H. Jerome Coyne III[1,4], Soumya De[1], Mark Okon[1], Sean M. Green[2,5], Niraja Bhachech[2], Barbara J. Graves[2,3] and Lawrence P. McIntosh[1*]

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

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

[3]Howard Hughes Medical Institute, Chevy Chase, MD, 20815, USA

[4]current address: Department of Chemistry, Indiana University, Bloomington, IN 47405, USA

[5]current address: Fred Hutchinson Cancer Research Center, Seattle, WA, 98109-1024, USA

* Corresponding author:
Lawrence P. McIntosh
Department of Biochemistry and Molecular Biology
Life Sciences Centre, 2350 Health Sciences Mall
University of British Columbia
Vancouver, B.C.
Canada, V6T 1Z3
Phone: (001) 604-822-3341
Fax: (001) 604-822-5227
E-mail: mcintosh@chem.ubc.ca

## Protein Expression and Purification for NMR Experiments

All ETV6 fragments used for the NMR experiments were expressed in frame with an N-terminal $His_6$-affinity tag and a thrombin cleavage site. The appropriate segments of the murine ETV6 gene (NP_031987) were cloned via PCR methods into a pET28b+ vector (Novagen) utilizing *Nde*I and *Hind*III restriction sites at the 5' and 3' ends, respectively.

The resulting plasmids were transformed into *E. coli* BL21 (λDE3) cells for protein expression. The pET28b+ vector provides kanamycin resistance and all media contained 70 mg/L of this antibiotic. In each case, an overnight culture in LB media was prepared using either a fresh transformant or a glycerol stock stored at -80 $^{o}$C. The 25 mL culture was pelleted by centrifugation at 3,000 rpm for 10 min, resuspended in sterile water, and used to inoculate 1 L of M9 media supplemented with 3 gm $^{13}C_6$-glucose and/or 1 gm of $^{15}NH_4Cl$ (Sigma-Aldrich or Cambridge Isotope Laboratories), as appropriate. This was grown in shaker flasks at 37 $^{o}$C, and when the $OD_{600}$ reached 0.6, protein expression was induced with 1 mM IPTG. After further growth at 30 $^{o}$C for 6-9 hours, the cells were collected by centrifugation in a Sorvall GSA rotor at 4,000 rpm, washed by resuspension with 0.25 M sucrose, split into two 50 mL disposable media tubes, spun again at 4,000 rpm, and stored at -80 $^{o}$C for future purification.

The cell pellet resuspended in 40 mL per tube of lysis buffer (100 mM Tris pH 8.2, 1 M NaCl, 20 mM imidizole, 5 mM tris(2-carboxyethyl)phosphine (TCEP), and 2-5 mM fresh phenylmethanesulfonylfluoride (PMSF) and stored on ice. Once thawed, the cells were disrupted by 5 cycles of sonication (1 min at full power and 1 min on cooling on ice) using a Branson Sonifer 250, followed by centrifugation for 30 min in a Sorvall SS34 rotor at 10,000 rpm. The clarified solution was loaded on to a 5 mL HP-FF $Ni^{+2}$-NTA resin column (GE Healthcare) and washed with 30 mL of wash buffer (100 mM Tris pH 8.2, 20 mM imidizole, 200 mM NaCl, 1 mM TCEP). The $His_6$-tagged protein was eluted using a 20-column volume gradient to 100% elution buffer (100 mM Tris pH 8.2, 1 M imidizole, 200 mM NaCl, 1 mM TCEP). The eluent was monitored at 280 nm while collecting 3 mL fractions. Fractions that exhibited a high absorbance were assessed for purity by SDS-Tricine gel electrophoresis. The highest purity

fractions were pooled, concentrated and exchanged by ultrafiltration with an Amicon 3 kDa MWCO concentrator into thrombin cleavage buffer (50 mM phosphate pH 8.0, 150 mM NaCl). The amount of protein was determined by UV absorbance, utilizing its predicted molar absorptivity (http://web.expasy.org/protparam/). The $His_6$-affinity tag was cleaved by overnight incubation with thrombin at room temperature with constant inversion. Affinity tag cleavage was deemed complete as determined by MALDI-TOF mass spectrometry, using 3,5-dimethoxy-4-hydroxycinnamic acid as the matrix. The resulting proteins, with four additional non-native N-terminal residues (Gly-Ser-His-Met), are denoted as $ETV6^{R426}$ (residues Arg335-Arg426), $ETV6^{R426'}$ (Gly329-Arg426), $ETV6^{Q436}$ (Arg335-Gln436), $ETV6^{D446'}$ (Gly329-Asp446), and $ETV6^{R458}$ (Arg335-Arg458). In the cases of $ETV6^{R426'}$ and $ETV6^{D446'}$, Cys334 was mutated to serine to prevent oxidation problems.

Once $His_6$-tag cleavage was confirmed, p-aminobenzamidine beads (3 mg protein : 250 μL beads) were added to remove the thrombin protease. The solution was mixed by inversion for at least 1 hr at room temperature, and the beads removed by centrifugation at 3,000 rpm for 10 min using a Beckman table top centrifuge. The solution was concentrated to 2-3 mL using an Amicon 3 kDa MWCO concentrator, and then loaded onto a Superdex 75 gel filtration column (2.6 cm diameter, 70 cm length; GE Healthcare). The protein sample was eluted with Gel Filtration Buffer (50 mM phosphate pH 5.8, 200 mM NaCl) at 2 mL/min. The eluent was monitored at 280 nm while collecting 2 mL fractions. The highest purity fractions, as assessed by SDS-tricine gel electrophoresis, were pooled and concentrated with buffer adjustment using an Amicon 3K Da MWCO concentrator. Final protein yields averaged about 6 mg per litre of M9 media, with $ETV6^{Q436}$ produced in the greatest amount. Towards the end of this study, it was found that yields could be increased ~2-fold by including 3M guanidinium chloride in the Lysis Buffer, and allowing refolding of the $Ni^{+2}$-NTA resin-bound protein during the wash step.

Unless stated otherwise, the NMR samples were prepared in a final volume of 300 - 600 μL at a concentration of 0.4 to 0.6 mM in sample buffer (50 mM phosphate, 200 mM NaCl, pH 5.8). To prevent any possible proteolytic degradation, 1 μL of EDTA Free Roche inhibitor

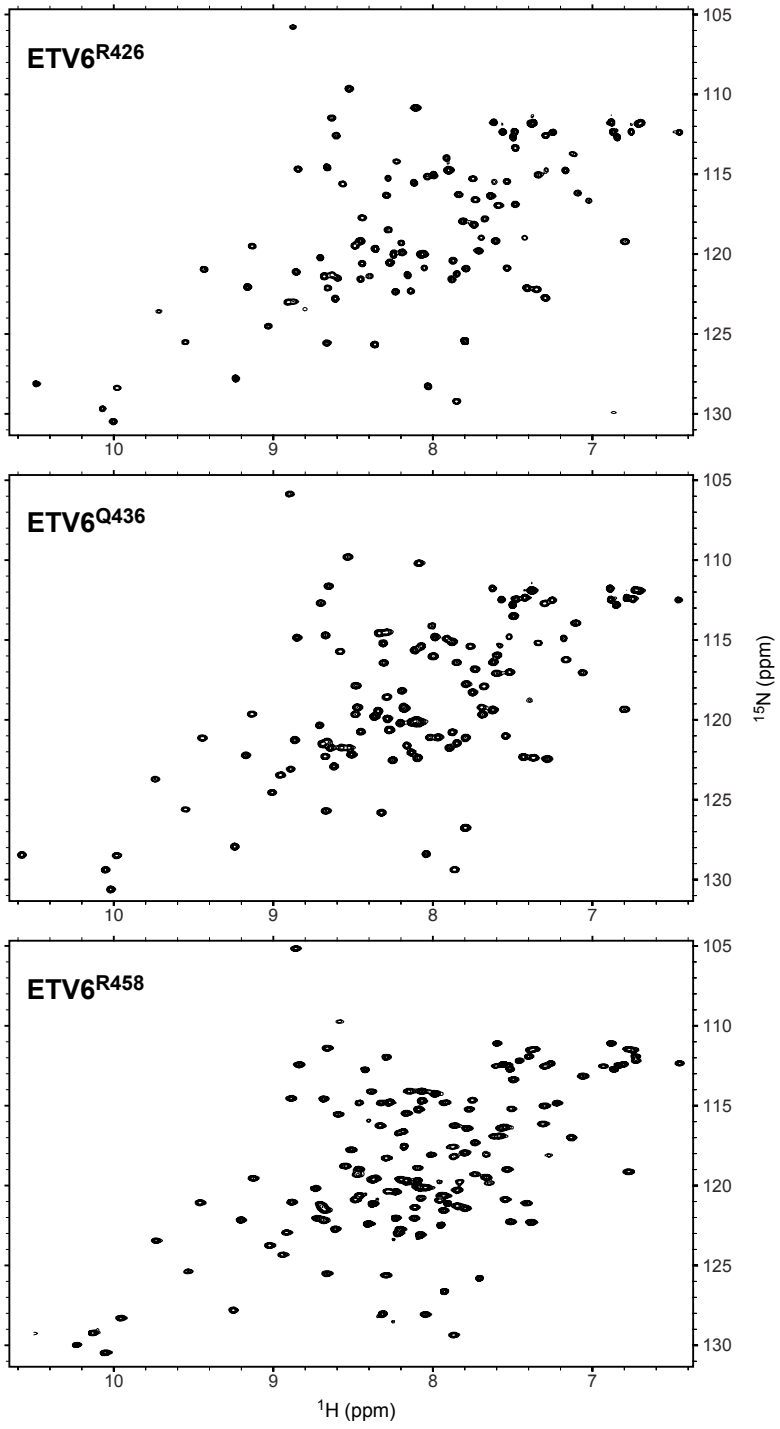
solution (prepared by dissolving 1 tablet in 1 mL of gel filtration buffer) was also added per 50 μL NMR sample.

### Protein Expression and Purification for EMSA Experiments

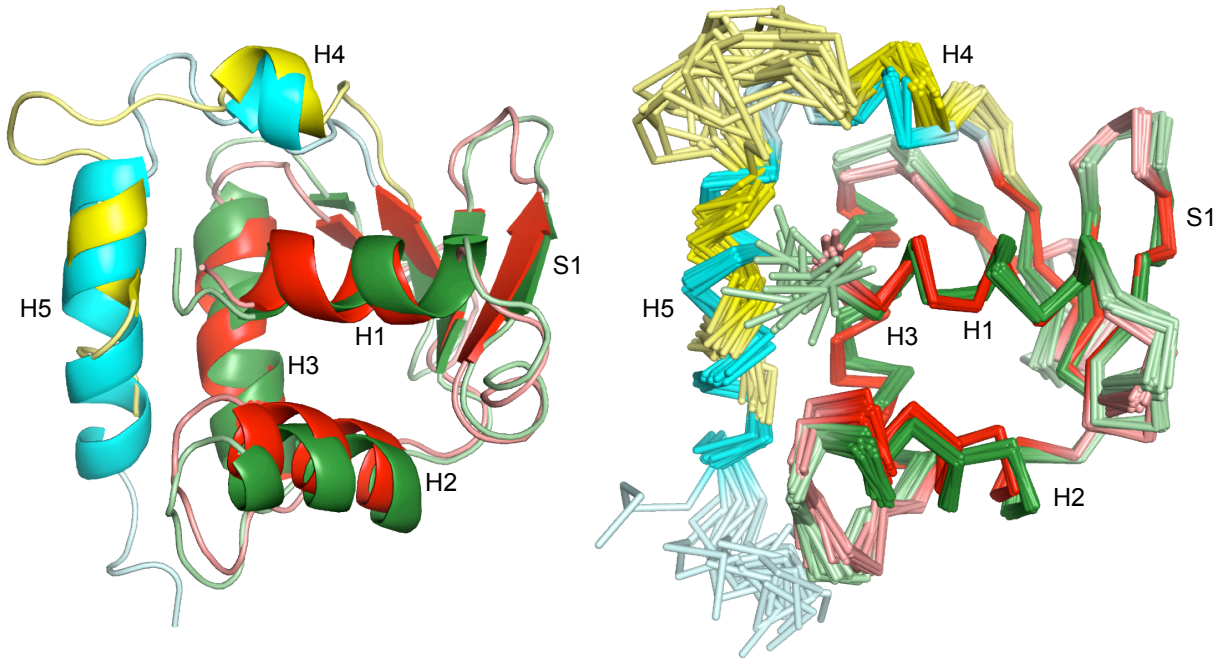
The plasmids were transformed into *E. coli* BL21 (λDE3) cells for protein expression. In each case, one liter of L-Broth, including kanamycin (100 μg/ml) for pET28b+ vector selection, was inoculated with an overnight culture. Cells were grown to an $OD_{600}$ of 0.8 and induced with 1 mM IPtG for 2 hours. The cell pellet was resuspended in 30 mL of cold 25 mM Tris-Cl pH 7.9, 1 M NaCl, 1 mM EDTA, 1 mM DTT, 1 mM phenylmethlysulfonyl fluoride, and 10 mM imidazole. This pellet was then lysed by sonication on ice and followed by centrifugation at 180,000 g for 45 min. The clarified supernatant was loaded onto four interlocked 5 mL HisTrap FF column (GE Healthcare) and washed with HisTrap Buffer (25 mM Tris-Cl pH 7.9, 1M NaCl, 0.1 mM EDTA, 1 mM BME) at 10 mM imidazole. The protein was eluted by a gradient up to 500 mM imidazole in HisTrap elution buffer. The eluent was monitored at 280 nm and the appropriate peak collected and dialyzed overnight at 4 $^{o}$C against Dialysis Buffer (20mM sodium citrate pH 5.3, 150 mM KCl, 1mM EDTA, 1mM DTT, 0.2mM PMSF and 10% glycerol). Simultaneously, the $His_6$-affinity tag was cleaved by adding 12 units of thrombin and adjusting the Dialysis Buffer to 2.5 mM CaCl. The protein was chromatographed through tandem DEAE resin and 5 mL S-Sepharose SPHP (GE healthcare) columns in the Dialysis Buffer. The ETV6, which flowed through the anion exchange column and bound the cation exchange column, was eluted by a linear gradient between 150 and 650 mM KCl in Dialysis Buffer. The ETV6 peak, monitored at 280 nm, eluted at approximately 500 mM KCl. After assessment by SDS-PAGE, the peak fractions were pooled then concentrated by Amicon 3kD MWCO filtration and loaded on a Superdex 75 column (2.6 cm diameter, 60 cm length; GE Healthcare) in buffer (20 mM sodium citrate pH 5.3, 500 mM KCl, 1 mM EDTA, 1 mM DTT, 0.2 mM PMSF and 10% glycerol). Purified protein, estimated >95% purity, was stored at -80 $^{o}$C in 30 μL aliquots for one-time use.

## Supplemental Figures

**Supplemental Fig. S1.** Sequence alignment of structurally characterized ETS domains (with Protein Data Bank accession codes). Capitalized residues are within the experimentally determined structure and used for alignment with Clustalw2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The arrows indicate the N- and C-terminal boundaries of ETV6$^{R426}$, ETV6$^{Q436}$, and ETV6$^{R458}$. Lower case grey residues are flanking sequences from the full length human (h) or mouse (m) protein and are not explicitly aligned. A period (.) indicates that the sequence continues and thus is not the true N- or C-terminus. Helices (red) and stands (orange) of the ETS domain and helices (dark cyan) of the appended regions were identified using the DSSP algorithm (http://2struc.cryst.bbk.ac.uk/). Residues shown in Supplemental Fig. S7 for ETV6, ETS1, and GABPA are identified with an asterisk (*). See also Supplemental Fig. S6 of Wei et al. (*EMBO J*. 29:2147-2160, 2010) for a more extensive alignment of ETS domain sequence from numerous ETS family members.

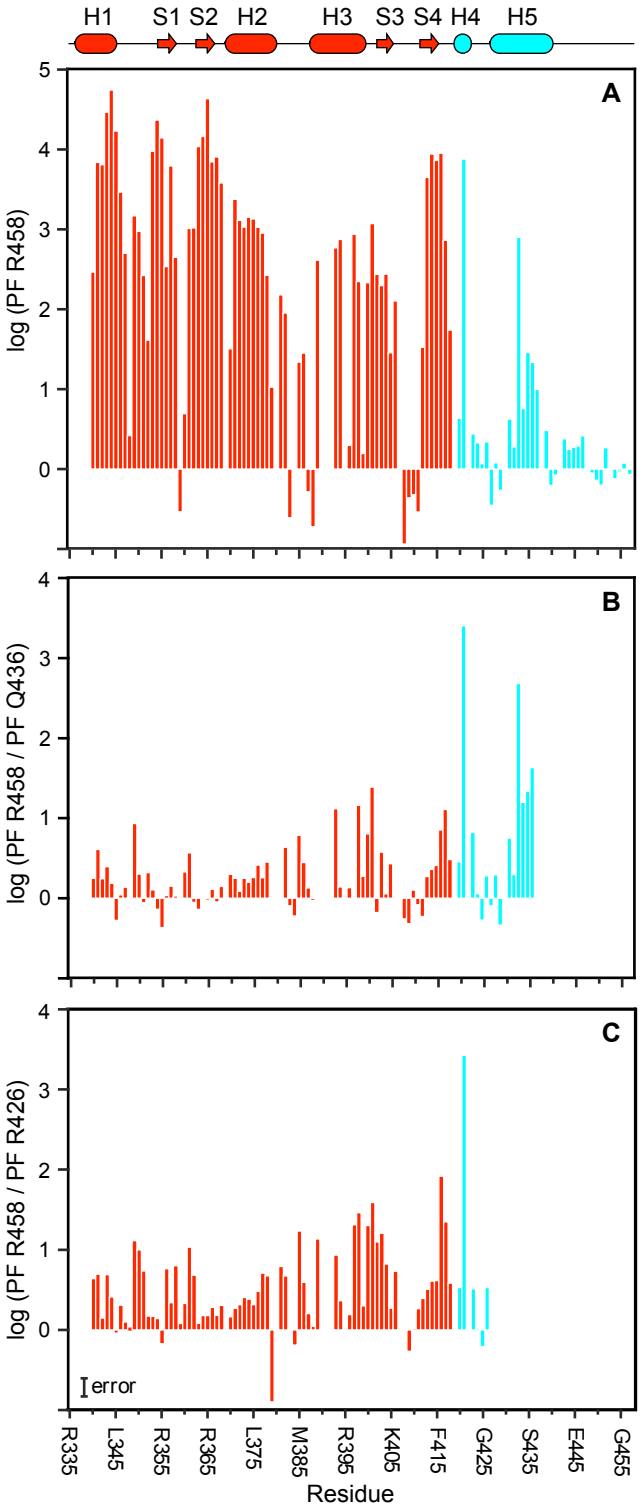



**Supplemental Fig. S2.** Sensitivity-enhanced $^{15}$N-HSQC spectra of the ETV6$^{R426}$, ETV6$^{Q436}$, and ETV6$^{R458}$ constructs (30 $^{o}$C, 50 mM phosphate, 200 mM NaCl, pH 5.8). The chemical shift assignments of the latter two have been deposited in the BioMagResBank under accession codes 17741 and 17742, respectively.



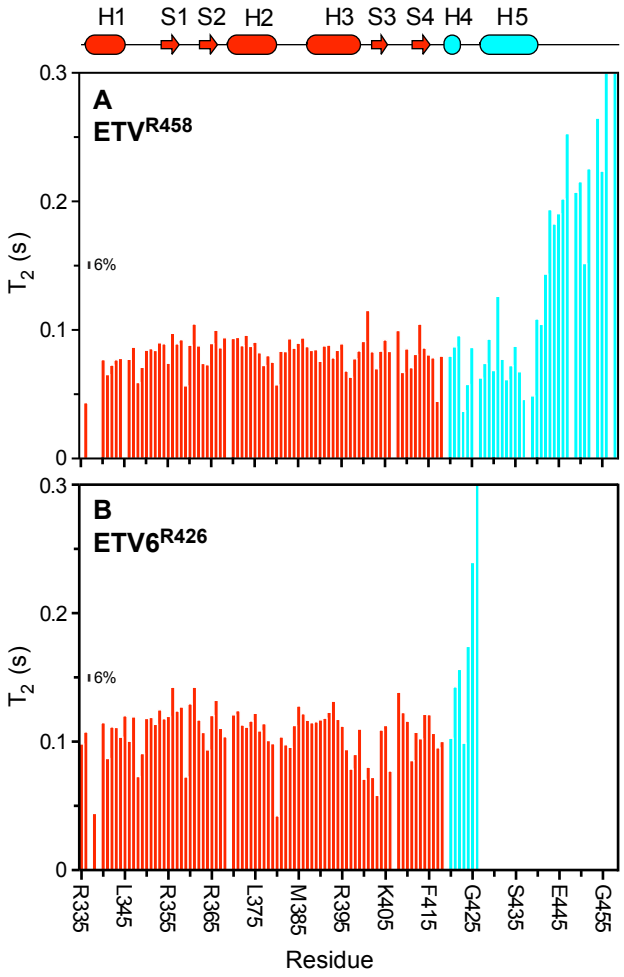

**Supplemental Fig. S3.** Superimposed cartoon and ribbon diagrams of the NMR-derived structural ensembles of murine ETV6$^{R458}$ (this study) and an undocumented fragment from the human ETV6 ortholog (2DAO.pdb; residues by 333-438 by murine numbering). The coordinates were aligned with PyMol based on helices H1, H2, and H3, and the secondary structural elements in the cartoon defined according to Supplemental Fig. S1. The ETS domain of ETV6$^{R458}$ is colored in red (helices and strands) and light red (coil regions), whereas the CID is in cyan (helices) and light blue (coils). The ETS domain of 2DAO is colored in green (helices and strands) and light green (coils), whereas the CID is in yellow (helices) and light yellow (coils). Disordered terminal residues have been removed for clarity.

The ETS domain coordinates of the two constructs align closely. However, in the 2DAO.pdb ensemble, the CID has a truncated helix H5 and higher rms deviations than the better defined ETS domain. Although translated in the vertical direction of the figure relative to ETV6$^{R458}$, helix H5 still clearly blocks the DNA binding interface of the human ETV6 construct.

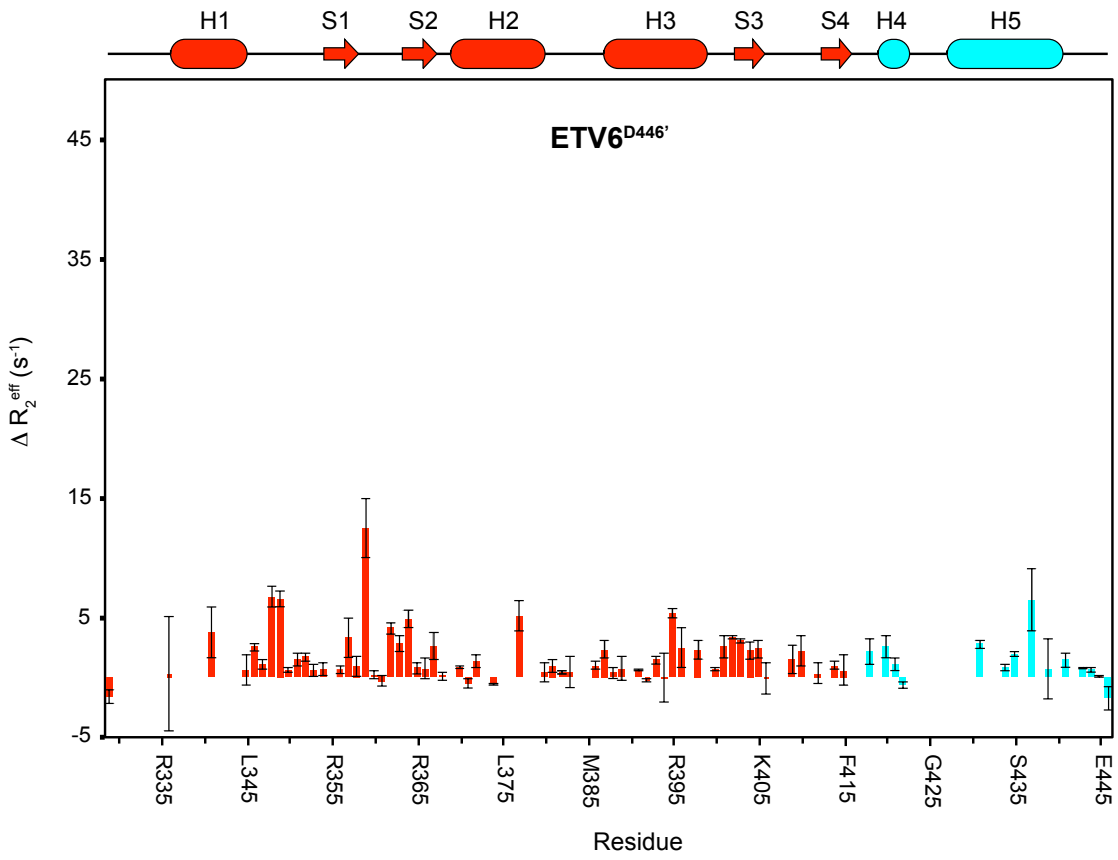



**Supplemental Fig. S4.** (A) HX protection factors (PF = $k_{pred}/k_{ex}$) for the ETS domain (red) and CID (cyan) of ETV6$^{R458}$ derived using $k_{ex}$ values measured at 30 $^{o}$C via either slow proton-deuterium or fast proton-proton exchange experiments. These data are reproduced from Fig. 4A.

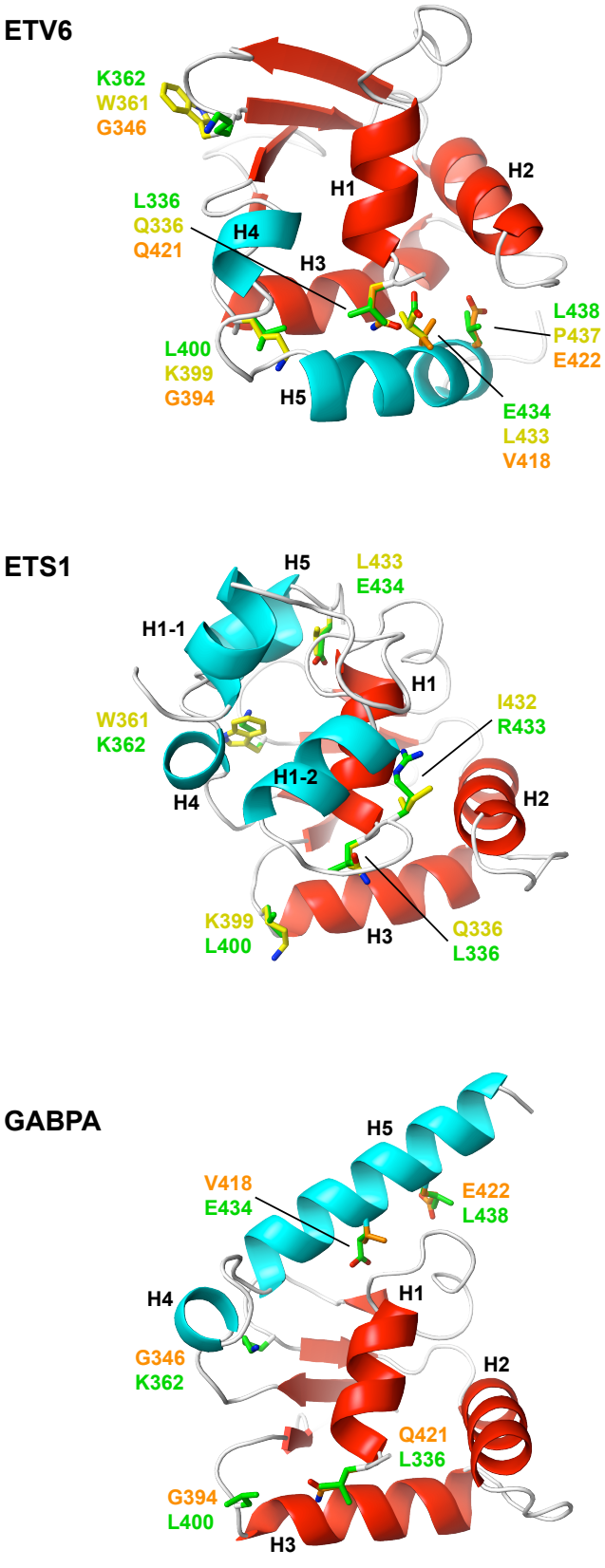
(B) The relative PF's for corresponding residues in ETV6$^{R458}$ versus ETV6$^{Q436}$. (C) The relative PF's for corresponding residues in ETV6$^{R458}$ versus ETV6$^{R426}$. Missing data points correspond to prolines, amides with overlapping signals, or amides for which HX rates could not be determined reliably by either exchange method (i.e., exchange lifetimes between several seconds and a few minutes). The indicated magnitude of the estimated PF or PF ratio errors, on a logarithmic scale, is based on assumed worse case errors of ±20% in both measured and predicted exchange rates. A cartoon representation of the secondary structure of ETV6$^{R458}$ is also shown with coloring as in Fig. 2.




**Supplemental Fig. S5.** Amide $^{15}N$ $T_2$ relaxation experiments reveal msec-μsec conformational exchange in the ETS domain that is dampened by the CID. Shown are $T_2$ values for the ETS domain (red) and CID (cyan) of (A) ETV6$^{R458}$ and (B) ETV6$^{R426}$. Unusually long $T_2$ lifetimes, as seen for the disordered C-terminal tails of the two proteins, reflect enhanced mobility on the nsec-psec timescale. In contrast, short $T_2$ lifetimes, as exhibited by residues spanning helix H3 to strand S3 in uninhibited ETV6$^{R426}$ but not inhibited ETV6$^{R458}$, result from conformational exchange broadening on the msec-μsec timescale. This exchange is best quantitated using the relaxation dispersion experiments presented in Fig. 6 and Supplemental Fig S6. Bars for Trp456 (0.31 $s^{-1}$) and Arg458 (0.49 $s^{-1}$) in ETV6$^{R458}$ and Arg426 (0.35 $s^{-1}$) in ETV6$^{R426}$ have been truncated, and the average fitting error is ± 3% (magnitude indicated). Missing data points correspond to prolines and amides with weak or overlapping signals. A cartoon representation of the secondary structure of ETV6$^{R458}$ is also given with coloring as in Fig. 2.




**Supplemental Fig. S6. Amide $^{15}$N relaxation dispersion experiments reveal that the msec-µsec conformational exchange exhibited by ETV6$^{R426'}$ is not detected in inhibited ETV6$^{D446'}$ due to the CID.** Plotted are $\Delta R_2^{eff}$ values for ETV6$^{D446'}$ from relaxation dispersion experiments with $\nu_{CPMG}$ of 50 Hz versus 1000 Hz, recorded using an 850 MHz NMR spectrometer. In contrast to uninhibited ETV6$^{R246'}$ with large $\Delta R_2^{eff}$ values (see Fig. 6A), ETV6$^{D446'}$ and ETV6$^{R458}$ (not shown) do not exhibit conformational exchange broadening. Thus, the CID dampens motions of the ETS domain in these latter fragments.




**Supplemental Fig. S7.** Modeling studies identify potential residues that establish the differential positions of helices H4 and H5 in ETS family members. Top: the experimentally-

determined structure of ETV6$^{R458}$ (this study) with key residues shown in stick format (carbon, green; oxygen, red; nitrogen, blue) along with the modeled corresponding side chains of ETS1 (carbon, yellow) and GABPA (carbon, orange). Middle: the structure of ETS1 (1R36.pdb) with key residues (yellow) shown with the modeled corresponding side chains of ETV6 (green). Bottom: the structure of GABPA (1AWC.pdb) with key residues (orange) shown with the modeled corresponding side chains of ETV6 (green). The helices and strands of the ETS domains are in red cartoon format, and the appended helices in cyan. In each case, Swiss-Model {Arnold, 2006 #149} was used to "mutate" the sequence of one ETS family member into the experimentally determined structure of the other member.

Note that the locations of the appended helices likely depend upon the combined positive, neutral, and negative effects of several amino acids within the context each possible structure, rather than on one signal lynchpin residue. That is, in any particular case, the amino acid differences may favor or disfavor one structure, whereas the reverse change may be relatively neutral. For example, substitution of the exposed Lys362 in the structure of ETV6$^{R458}$ with the corresponding tryptophan of ETS1 or glycine of GABPA would likely be tolerated, but the reciprocal substitution of the buried Trp361 in ETS1 with a lysine of ETV6 would be not be favorable. Thus, the sequence at this particular position would disfavor ETV6 from adopting the ETS1 fold, but not visa versa. However, changes at other positions would disfavor ETS1 from folding as ETV6.