Supporting Information

Regan et al. 10.1073/pnas.1301726110

SI Materials and Methods

Protein Expression and Purification. Human *ERG* cDNA constructs were cloned into the modified ampicillin-resistant pET-based vector pHis-Parallel, which contains an N-terminal 6× Histidine tag and a protease site to allow cleavage by recombinant Tobacco Etch Virus protease (rTEV). All ERG protein constructs were expressed in *Escherichia coli* ArcticExpress (DE3) RIL cells (Stratagene) at 10 °C for 18 h with 1 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG). Isotopically labeled protein samples that would be used for NMR studies were grown in minimal media supplemented with 1 g/L ^{15}N-$(NH_4)_2SO_4$ and 5 g/L ^{12}C-D-glucose and 1% ^{15}N-Bioexpress (Cambridge Isotope) for uniformly ^{15}N-labeled samples, or 1 g/L ^{15}N-$(NH_4)_2SO_4$ and 2 g/L ^{13}C-D-glucose and 1% ^{15}N/^{13}C -Bioexpress for uniformly ^{15}N/^{13}C-labeled samples and induced as above. Protein samples to be used for all other experiments were grown in Terrific Broth (Fisher) supplemented with 4 mL/L glycerol.

Selenium-labeled ERGi protein samples were grown in minimal media as described (1), with the following modifications: All amino acids were initially present at 0.1 g/L except for methionine. Before induction, 62.5 mg/L of selenomethionine (Anatrace) and 100 nM cyanocobalamin (vitamin B_{12}) were added along with 1 g/L of all amino acids except methionine. Cells were induced at 30 °C for 4 h with 1 mM IPTG before harvesting and purification as described above. MALDI-TOF mass spectrometry analysis showed incorporation of >85% of selenomethionine at four of four potential sites and >95% incorporation into at least three of the four sites.

After induction, cells were harvested and lysed by passage through a cell disruptor and spun at high speed to produce a clarified lysate. The protein was purified on Nickel-NTA resin and washed thoroughly with 1 M KCl to remove contaminating DNA fragments, as measured by absorbance at 260 nm. The Histidine tag was removed by overnight cleavage with rTEV and passage down the Ni-NTA column a second time followed by a round of anion-exchange chromatography on SP Sepharose (GE Life Sciences). As a final polishing step for crystallography samples, ERG protein underwent size-exclusion chromatography. Purified protein was dialyzed to 0.5 M KCl, 25 mM Bis·Tris (pH 6.0), 5 mM DTT, and 10% (vol/vol) glycerol before storage at −80 °C.

Oligonucleotides. Oligonucleotides were based on those found to bind with high affinity to a protein closely related to ERG, Fli1 (2). The 16-mer for which ERG showed high affinity consisted of the sequence AGGACC**GGAA**GTAACT and its reverse complement, with the canonical Ets-protein binding site highlighted in bold. Lyophilized oligomers (Operon) were suspended in 10 mM Hepes at pH 7.5 and 10 mM KCl, and forward and reverse strands were combined 1:1 before being heated to 95 °C for 20 min, followed by gradual cooling to room temperature. Excess single-stranded oligomers were removed by Q-Sepharose (GE Life Sciences) anion-exchange chromatography. An identical protocol was used for oligonucleotides used in crystallization, but because of poor results, this fragment was shortened to a 12-mer consisting of the sequence GACC**GGAA**GTGG (and reverse complement). Appropriate fractions containing double-stranded DNA were pooled and concentrated to 1 mM before being frozen at −80 °C for storage.

Isothermal Titration Calorimetry. Isothermal titration calorimetry (ITC) experiments were conducted on a VP-ITC MicroCalorimeter (MicroCal) at 17 °C. All samples were concentrated to the appropriate concentrations before dialysis in the same buffer. All protein samples were concentrated to ∼15 μM, and all DNA samples were concentrated to 200 μM. Immediately before the experiment, all samples were thoroughly degassed and cooled to temperature. Experiments were run by using a series of 35 injections of DNA into ERG protein. Evolved heats were referenced against degassed ddH_2O, and baseline correction was achieved by manual fitting and subtraction of residual heats of dilution. The data were processed by using the Origin-based MicroCal ITC-customized software and fit by using a one-site binding model.

Crystallization. NMR-based studies identified stable fragments of ERG suitable for crystallization. After purification as described above and dialysis to 25 mM Tris at pH 7.5, 1 mM DTT, and either 0.5 M KCl (for ERG protein alone) or no salt (for the ERG:DNA complex), the protein was concentrated to 3 mg/mL to be used in sitting-drop crystal screens. Within 5–7 d, crystals formed in wells with ERGi or ERG(289–388) protein solution combined 1:1 with well solution containing 1.4 M sodium citrate at pH 6.5. ERGi crystals grew to a maximum size of ∼200 μm × 50 μm × 50 μm in 2–3 wk at 17 °C, whereas ERG(289–388) crystals grew much larger, to ∼800 μm × 200 μm × 200 μm. The ERGu construct consisting of only the minimal Ets domain did not form crystals. To generate crystals of the ERGi:DNA complex, protein and DNA were combined in a 1:1 molar ratio and concentrated to 12 mg/mL; after 3–4 mo, exceedingly fragile rhombic plates formed with dimensions of 300 μm × 150 μm × 10 μm in 0.2 M NH_4Cl and 20% (wt/vol) PEG 3350. Before freezing in liquid nitrogen for transport, all crystals were cryoprotected by soaking in mother liquor with 50% (vol/vol) glycerol.

Structure Determination and Refinement. Initial phases were generated from single anomalous diffraction data of SeMet-ERGi crystals gathered at 0.9787 Å at 100 K; this structure was then used as a model for molecular replacement to solve the structure of native ERG(289–388); native crystal datasets were collected at 1.00 Å. To solve the ERGi:DNA complex, we used a molecular replacement ensemble containing the ERGi model and the previously determined structure of the c-fos promoter DNA (1BC7) (3). Phases were determined by using Phaser, and finished structures were validated with MolProbity (4). Of all residues in ERG(289–388), 100% were in Ramachandran favored/allowed regions. In both ERGi and the ERGi:DNA complex, 99% were allowed and 1% of residues were outliers. Figures were prepared by using PyMOL (5).

NMR Spectroscopy. All NMR-based experiments were acquired in buffer containing 200 mM $MgSO_4$, 20 mM KP_i, 5 mM DTT, 0.01% sodium azide, and 5% D_2O at a final pH of 6.0. A high ionic strength was necessary to keep ERG stable in solution, and a $MgSO_4$ solution allowed for a substantially shorter 90° ^{1}H pulse than required in equivalent solutions of KCl, as observed (6). In the presence of DNA for experiments involving the ERG:DNA complex, high salt was unnecessary and the experimental buffer was made up without $MgSO_4$ and a final pH of 6.9. All experiments were recorded on samples concentrated to 0.6 mM protein at 30 °C. For experiments involving ERG complexed with DNA, 0.6 mM protein was combined with DNA at a molar ratio of 1:1.05, protein:DNA.

Peptide backbone assignments were made by using a set of multidimensional experiments on uniformly $^{15}N/^{13}C$-labeled samples. These experiments included HNCACB, CBCA(CO)NH, and HNCO, and to assist with limitations due to peak broadening from conformational exchange intermediate on the nitrogen timescale, we also performed an HCACO experiment. All experiments were performed on either a Varian 600 MHz magnet equipped with a Cryoprobe or Bruker 600 MHz or 800 MHz magnets equipped with ColdProbes. Rotational correlation times were estimated at 600 MHz by using the TRACT experiment as described (7), and subsequently more accurate values were determined from a trimmed mean (excluding the highest and lowest 10% of values) of R_2/R_1 relaxation rates as described (8). ^{15}N-Heteronuclear NOE experiments were performed on uniformly ^{15}N-labeled samples for both ERG alone and ERG in complex with DNA. In both cases, a relaxation delay of 3 s was used. T_1 experiments were conducted by using a series of relaxation delays of 2.1, 1.7, 1.3, 0.9, 0.7, 0.5, 0.3, 0.2, 0.1, and 0.01 s. T_2 experiments used relaxation delays of 10, 30, 50, 70, 90, 110, 130, 170, 210, and 250 ms. All three sets of experiments were performed at 600 MHz (14.1 T). Peak intensities were fit, and R_1 and R_2 relaxation rates were calculated by using SPARKY (T. D. Goddard and D. G. Kneller, University of California, San Francisco). Relaxation dispersion experiments were conducted at 18.8 T by using the pulse sequence described by Long et al. (9) with ν_{CPMG} frequencies of 80, 200, 280, 440, 600, 760, 840, and 920 Hz with a constant delay of 50 ms. The data were fit to a two-state model by using the program Nessy (10).

Electron Paramagnetic Resonance Spectroscopy. Wild-type ERG contains one cysteine at position 312 in the Ets domain, which was mutated to serine, whereas single sites chosen to be spin labeled were mutated to cysteine. These samples were prepared as described above, with the following modifications. After the initial Nickel-column purification step, rather than removing the His tag, the protein was concentrated and thoroughly degassed before desalting on a PD10 column (GE Healthcare) into a buffer containing 10 mM DTT. The sample was allowed to reduce before being desalted a second time into the above buffer with no DTT. To this solution was added a DMSO stock solution of (1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-methyl) methanethiosulfonate (Toronto Research Chemicals) (11, 12). The labeling process was allowed to proceed overnight at 4 °C before the sample was centrifuged to remove precipitation and loaded onto an SP-Sepharose cation-exchange column and extensively washed with degassed buffer to remove any residual free spin label. The protein was eluted, concentrated to ∼50 μM, and separated into individual 500-μL aliquots (12).

Approximately 50 μL of NTA-Nickel beads for each of the experimental conditions were resuspended in 100 μL of spin-labeled ERG and centrifuged. The supernatant was removed, and the beads were resuspended in fresh protein solution four more times to ensure that the Nickel beads were sufficiently bound with protein. Finally, the beads were washed five times with 100 μL each of buffer containing 300 mM KCl, 25 mM Tris at pH 7.5, 10 mM Imidazole, and any of the multiple concentrations of sucrose (10–40%, wt/vol) or 100 μM DNA, as indicated.

A slurry of ∼50% Nickel bead-bound ERG was loaded into 0.6 mm ID × 0.84 mm OD glass capillary tubes (VitroCom) for CW-based EPR experiments with an incident microwave power of 2 mW at room temperature.

1. Tong KI, Yamamoto M, Tanaka T (2008) A simple method for amino acid selective isotope labeling of recombinant proteins in E. coli. *J Biomol NMR* 42(1):59–67.
2. Mao X, Miesfeldt S, Yang H, Leiden JM, Thompson CB (1994) The FLI-1 and chimeric EWS-FLI-1 oncoproteins display similar DNA binding specificities. *J Biol Chem* 269(27): 18216–18222.
3. Mo Y, Vaessen B, Johnston K, Marmorstein R (1998) Structures of SAP-1 bound to DNA targets from the E74 and c-fos promoters: Insights into DNA sequence discrimination by Ets proteins. *Mol Cell* 2(2):201–212.
4. Chen VB, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12–21.
5. Baugh EH, Lyskov S, Weitzner BD, Gray JJ (2011) Real-time PyMOL visualization for Rosetta and PyRosetta. *PLoS ONE* 6(8):e21931.
6. Wolf-Watz M, Xie XQ, Holm M, Grundström T, Härd T (1999) Solution properties of the free and DNA-bound Runt domain of AML1. *Eur J Biochem* 261(1):251–260.
7. Lee D, Hilty C, Wider G, Wüthrich K (2006) Effective rotational correlation times of proteins from NMR relaxation interference. *J Magn Reson* 178(1):72–76.
8. Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: Application to staphylococcal nuclease. *Biochemistry* 28(23):8972–8979.
9. Long D, Liu M, Yang D (2008) Accurately probing slow motions on millisecond timescales with a robust NMR relaxation experiment. *J Am Chem Soc* 130(8): 2432–2433.
10. Bieri M, Gooley PR (2011) Automated NMR relaxation dispersion data analysis using NESSY. *BMC Bioinformatics* 12:421.
11. Battiste JL, Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* 39(18): 5355–5365.
12. Liang B, Bushweller JH, Tamm LK (2006) Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy. *J Am Chem Soc* 128(13):4389–4397.

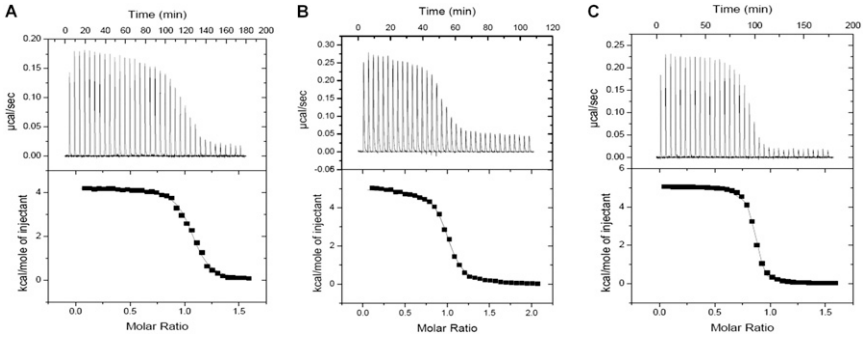

Fig. S1. Selected representative ITC isotherms of full-length ERG (*A*), autoinhibited ERG (*B*), and the minimal Ets domain binding to DNA (*C*).

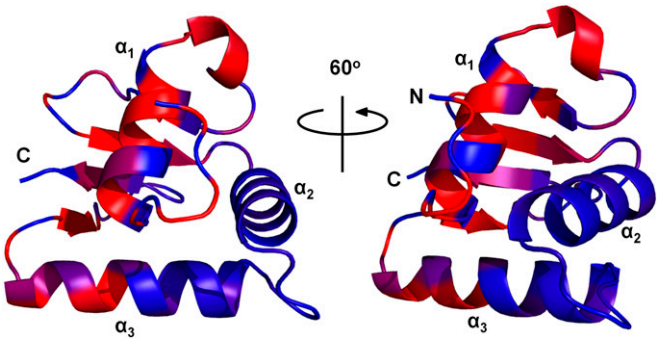

Fig. S2. Chemical shift perturbations mapped onto a ribbon representation of ERGu. Residues undergoing large (>0.1 ppm) changes (red) cluster primarily on one face of the protein where the autoinhibitory cassette docks against the Ets domain.

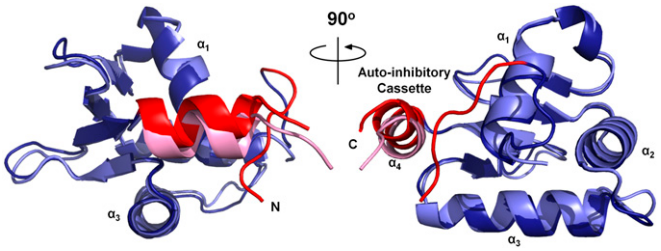

Fig. S3. Comparison of DNA-free ERGi (with Ets domain in dark blue and autoinhibitory cassette in red) and DNA-bound ERGi (light blue and pink, respectively) demonstrating the shift undergone in the CID upon DNA binding.

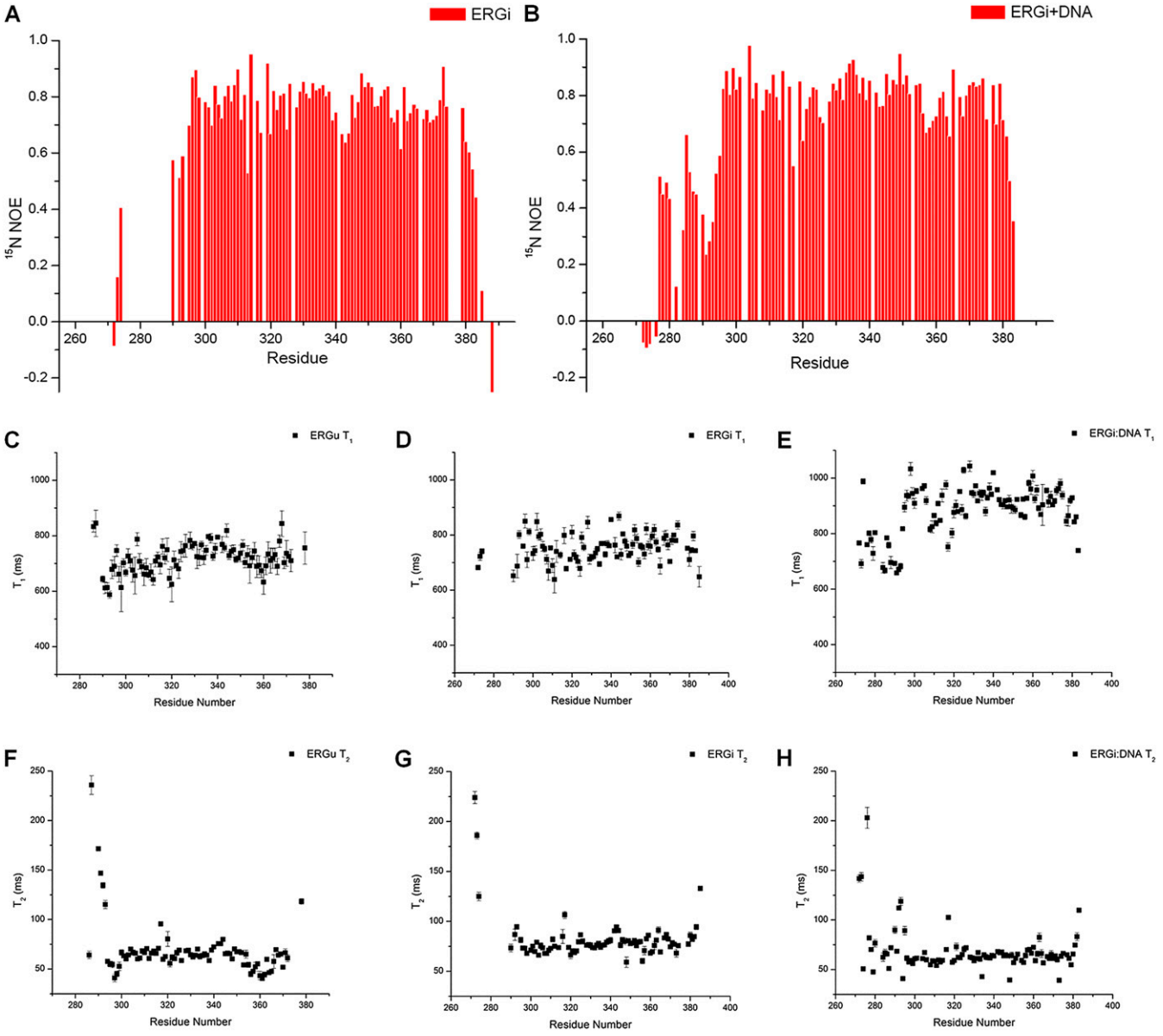

Fig. S4. ^{15}N-^{1}H heteronuclear NOE values for ERGi (*A*) and ERGi bound to DNA (*B*), measured as backbone amide peak intensities after 3-s radio frequency saturation vs. no saturation. Although we could not make full assignments of the NID of unbound ERGi in solution because of the absence of these peaks in the spectrum and, therefore, could not collect nuclear Overhauser effect (hetNOE) data for these residues, it is interesting to note that when bound to DNA, these residues show low, but not negative, NOE ratios, consistent with a high degree of conformational flexibility. (*C*) ERGu ^{15}N T_1 relaxation rates for each assigned residue in the protein. (*D*) ERGi ^{15}N T_1 relaxation rates (*E*) ERGi:DNA complex ^{15}N T_1 relaxation rates. (*F*) ERGu ^{15}N T_2 relaxation rates. (*G*) ERGi ^{15}N T_2 relaxation rates. (*H*) ERGi:DNA complex ^{15}N T_2 relaxation rates.

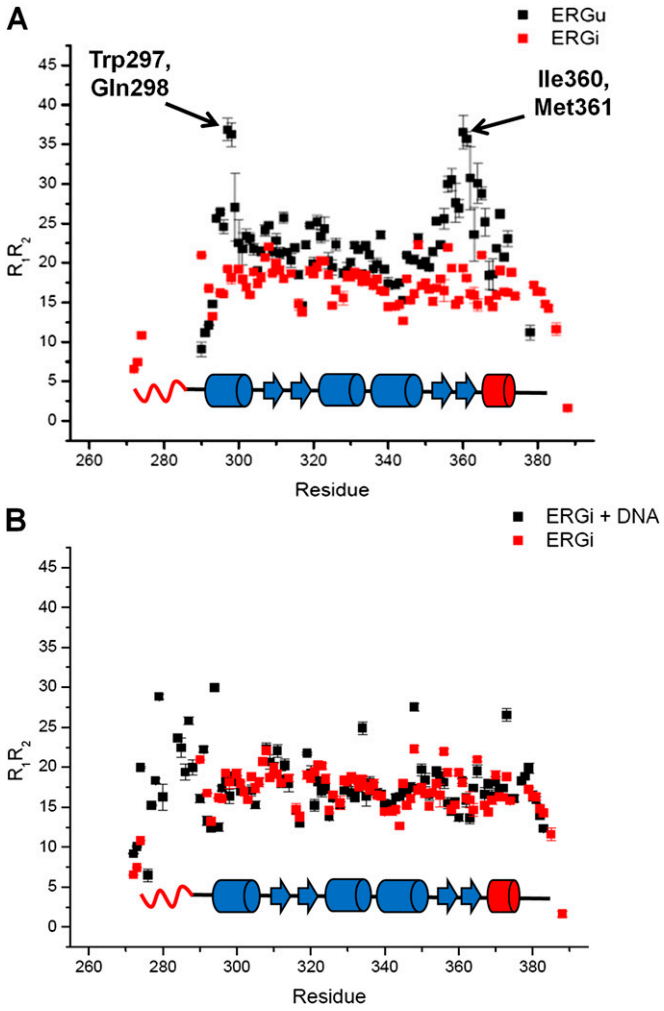

Fig. S5. Backbone dynamics of uninhibited, autoinhibited, and DNA-bound ERG at 14.1 T. (*A*) Overlay of R_1R_2 values for ERGi (red) and ERGu (black), with secondary structural elements of ERGi illustrated below. Autoinhibitory regions are in red. Key residues Trp297/Gln298 and Ile360/Met361 display elevated R_1R_2 values in ERGu. Error bars represent ±SE. (*B*) Overlay of R_1R_2 values for ERGi (red) and ERGi bound to DNA (black). The gap in N-terminal values for ERGi represents those residues that are in conformational exchange and could not be assigned in the absence of DNA.

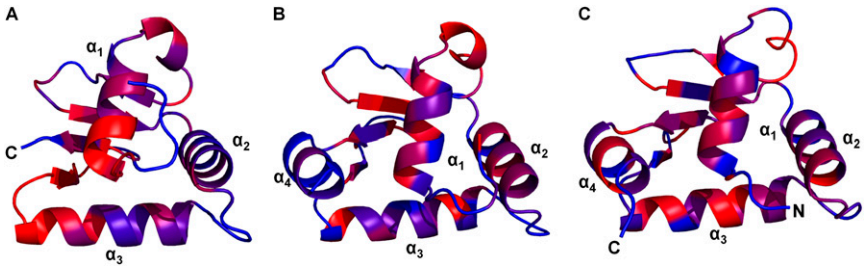

Fig. S6. Regions of increased dynamics in ERG. R_1R_2 values mapped onto the structures of ERGu (*A*), ERGi (*B*), and ERGi (*C*) when bound to DNA. Elevated values are shown in red; reduced values are shown in blue.

Table S1. Complete ITC results

Construct	n	$-T\Delta S$	ΔH	ΔG	K_D	Fold inhibition
1–462	1.05 (±0.04)	−13,369 (±29)	4,205 (±30)	−9,164 (±42)	121 (±5)	3.3
210–462	1.10 (±0.02)	−13,572 (±87)	4,303 (±16)	−9,269 (±88)	104 (±5)	2.8
272–412	0.94 (±0.03)	−14,181 (±232)	4,900 (±25)	−9,281 (±233)	101 (±8)	2.7
272–388	0.98 (±0.03)	−13,891 (±46)	4,924 (±24)	−8,967 (±52)	176 (±10)	4.8
289–412	0.98 (±0.06)	−14,732 (±116)	5,062 (±32)	−9,670 (±120)	51 (±10)	1.4
272–378	0.99 (±0.03)	−14,645 (±29)	4,731 (±18)	−9,914 (±34)	33 (±3)	0.9
289–388	0.99 (±0.02)	−14,819 (±29)	5,024 (±25)	−9,795 (±38)	41 (±4)	1.1
289–378	1.04 (±0.02)	−14,703 (±145)	4,851 (±24)	−9,852 (±147)	37 (±3)	1.0
ERGu Y354F	1.00 (±0.03)	−14,761 (±174)	5,503 (±24)	−9,258 (±176)	106 (±6)	2.9
ERGi Y354F	1.01 (±0.04)	−15,457 (±74)	6,565 (±34)	−8,892 (±81)	200 (±13)	5.4
ERGi S283A	1.01 (±0.04)	−15,269 (±334)	5,968 (±43)	−9,301 (±336)	97 (±11)	2.6

Isothermal titration calorimetry values for the ERG truncation constructs spanning the given residues. ΔG was calculated by using experimental values for $T\Delta S$ and ΔH in cal/mol, K_D is listed in nanomolar, and ±SE for triplicate experiments is listed in parentheses.

Table S2. ERG R_{Ex} values from relaxation dispersion

Residue	ERGu	ERGu+DNA (20:1)	ERGi	ERGi+DNA (1:1.05)
272	—	—	—	—
273	—	—	—	—
274	—	—	—	—
275	—	—	—	—
276	—	—	—	—
277	—	—	—	—
278	—	—	—	—
279	—	—	—	—
280	—	—	—	—
281	—	—	—	—
282	—	—	—	—
283	—	—	—	—
284	—	—	—	—
285	—	—	—	17.89
286	—	—	—	—
287	—	—	—	—
288	—	—	—	—
290	18.99	12.33	23.63	—
291	15.57	9.12	—	—
292	16.05	13.45	2.28	—
293	12.47	10.76	13.22	—
294	14.73	7.83	—	—
295	15.54	13.82	—	—
296	18.25	16.63	—	—
297	10.54	18.81	10.69	—
298	—	—	—	—
299	11.71	12.99	—	—
300	9.41	10.41	4.94	—
301	9.62	11.13	7.25	—
302	18.96	20.66	10.38	—
303	12.10	10.63	12.46	—
304	6.65	—	11.22	—
305	11.05	9.22	—	—
306	13.41	—	5.20	—
307	16.15	19.86	10.55	—
308	12.02	7.08	4.48	—
309	6.02	3.49	6.74	—
310	8.41	6.41	—	—
311	17.25	5.38	—	—
312	8.95	6.78	—	—
313	6.52	6.67	—	—
314	8.28	7.96	—	—
315	18.24	12.13	—	—
316	16.34	15.84	4.68	—
317	8.81	7.03	—	—
318	17.05	17.80	—	—
319	16.20	23.94	5.60	—
320	—	—	8.48	14.94
321	12.78	—	—	—
322	15.15	8.95	—	—
323	17.04	5.88	4.18	—
324	—	—	4.81	—
325	7.70	—	—	—
326	14.91	—	—	—
327	—	—	—	—
328	9.13	—	25.85	—
329	16.00	8.23	—	—
330	9.32	—	—	—
331	11.45	—	—	—
332	10.12	—	—	—
333	—	—	—	—
334	—	6.11	—	—
335	17.14	—	—	—

Table S2. Cont.

Residue	ERGu	ERGu+DNA (20:1)	ERGi	ERGi+DNA (1:1.05)
336	11.26	—	—	—
337	—	—	—	—
338	20.45	13.49	—	—
339	7.93	—	—	—
340	16.31	4.93	—	—
341	—	—	—	—
342	6.75	7.85	—	—
343	—	—	—	—
344	8.72	—	—	—
345	4.69	7.09	—	—
346	15.50	5.66	—	—
347	9.60	—	—	—
348	10.52	9.11	—	—
349	9.98	—	—	—
350	14.07	7.98	—	7.54
351	8.15	7.25	—	6.07
352	8.88	6.13	—	—
353	24.26	21.07	7.70	—
354	12.33	6.17	4.30	14.77
355	17.12	6.55	—	30.55
356	17.42	16.40	14.19	—
357	22.34	12.81	17.29	—
358	14.97	11.19	4.75	—
359	19.01	25.81	3.32	—
360	—	—	12.26	—
361	—	—	6.13	—
362	19.09	—	7.41	—
363	16.77	13.70	4.00	—
364	12.67	21.67	—	—
365	9.99	45.84	—	—
366	20.04	34.37	—	—
367	9.55	6.95	—	—
368	7.31	5.86	18.38	6.94
369	13.44	14.02	—	—
370	8.68	14.04	—	—
371	10.61	9.99	—	—
372	9.96	13.33	—	—
373	—	—	4.62	—
374	—	—	—	—
375	—	—	—	—
376	—	—	—	—
377	—	—	—	—
378	—	—	—	—
379	—	—	5.60	—
380	—	—	10.03	—
381	—	—	7.58	—
382	—	—	14.10	—
383	—	—	9.90	—

R_{Ex} values for all assigned residues in each ERG construct were calculated from peak heights in ^{15}N–^{1}H-NMR spectra collected at a heteronuclear frequency of 81.1 MHz. Values were calculated with the program NESSY by using a two-state exchange model.

Table S3. Data collection and refinement statistics

	ERGi:DNA	ERGi	ERG(289–388)
Space group	$P4_122$	$P6_522$	$P6_522$
Unit cell dimensions	$a = b = 51.47$, $c = 141.36$	$a = b = 44.74$, $c = 175.08$	a = b=44.70, c = 175.11
	$\alpha = \beta = \gamma = 90$	$\alpha = \beta = 90$, $\gamma = 120$	$\alpha = \beta = 90$, $\gamma = 120$
X-ray source	APS 22-ID	APS 22-ID	APS 22-ID
Solution method	MR (ERGi/1BC7)	SAD (3 SeMet Sites)	MR (ERGi)
Wavelength, Å	1.000	0.9787	1.000
Resolution range, Å	34.76–2.80	32.28–2.10	35.43–1.70
(Outer shell) , Å	(2.85–2.80)	(2.14–2.10)	(1.73–1.70)
Observations/unique	58,403/5,250	107,803/6,372	234,398/12,118
Completeness, %	99.13 (94.62)	94.77 (71.38)	99.27 (97.04)
Reflections R_{sym}, %	19.3 (52.2)	12.5 (48.0)	12.2 (39.8)
<I/σI>	14.20 (3.94)	13.57 (4.58)	23.45 (11.34)
<Redundancy>	11.0 (7.3)	16.3 (9.9)	19.3 (16.6)
Protein/DNA/solvent atoms	770/484/17	820/–/26	807/–/46
R_{work}, %	19.05 (26.44)	18.07 (20.42)	18.81 (21.72)
R_{free}, %	20.99 (34.71)	21.50 (28.98)	22.76 (30.92)
Ramachandran plot, most favored/additional, %	99/1.1	98/1	100/0
rmsd bond length, Å	0.015	0.022	0.022
rmsd dihedral, °	1.69	1.76	2.00

Data were collected on a single crystal for each structure. Values in parentheses are for outermost shell.