Nucleic Acids Research, 2025, **53**, gkaf205
https://doi.org/10.1093/nar/gkaf205
Molecular and Structural Biology

Structure and cooperative formation of a FLI1 filament on contiguous GGAA DNA sites

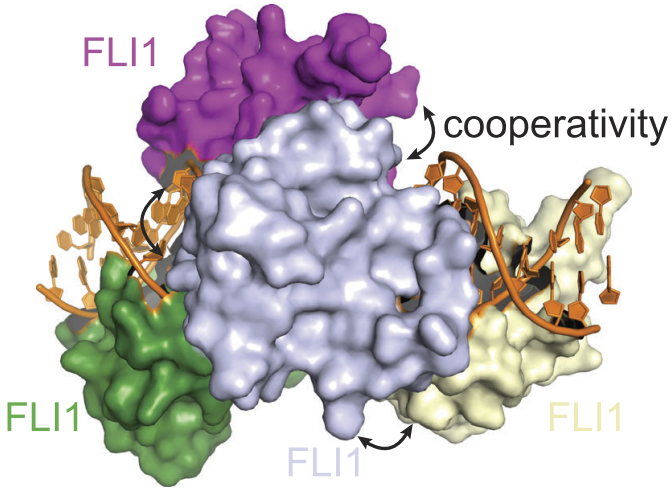
Caixia Hou[1] **and Oleg V. Tsodikov**[1,2,*]

[1]Department of Pharmaceutical Sciences, College of Pharmacy, University of Kentucky, Lexington, KY 40536, United States
[2]Markey Cancer Center, University of Kentucky, Lexington, KY 40536, United States
[*]To whom correspondence should be addressed. Email: oleg.tsodikov@uky.edu

Abstract

Ewing sarcoma, a pediatric cancer of bone and soft tissue, is driven in most cases by an abnormal oncogenic fusion of the N-terminal region of EWS with the C-terminal region of FLI1 (EWS–FLI1). The FLI1 region contains a conserved DNA-binding domain (DBD) essential for the oncogenesis. Binding of EWS–FLI1 to microsatellites composed of contiguous GGAA sites, shown previously to be critical for the oncogenic program of this fusion, is not well understood. In this study, we demonstrate that the FLI1 DBD binds cooperatively to contiguous GGAA sites, thereby forming a nucleoprotein filament. A series of crystal structures of two, three, and four FLI1 DBD proteins in complexes with DNA oligomers containing two, three, and four contiguous GGAA sites, respectively, reveal the structure of this filament and the basis for its cooperative formation. We expect this mechanistic insight to be an important milestone in our understanding of the oncogenic function of EWS-FLI1 and exploiting it as a drug target.

Graphical abstract



Introduction

Ewing sarcoma (ES) is an aggressive pediatric cancer of bone and soft tissue. Initial treatment of ES with a nontargeted drug combination, together with surgery and radiotherapy, result in the 5-year survival rate of ∼70%, but for those whose cancer recurs there are currently no efficacious therapeutic options, and the prognosis is poor [1]. Therefore, mechanistic knowledge of this and similar cancers is sorely needed to develop precision therapeutics. ES is characterized and driven by an abnormal fusion of the N-terminal region of EWS or FUS protein and the C-terminal region of one of five ETS transcription factors (FLI1, ERG, FEV, ETV1, or ETV4). The chromosomal translocation resulting in this fusion is the only common mutation and, in many cases, the only somatic mutation in this cancer [2, 3]. Fusions involving EWS are by far most common, to FLI1 in 85% of ES cases, to ERG in 10% and to one of the three other ETS family members (FEV, ETV1, and ETV4) in most of the remaining cases [4]. Fusions involving FUS are estimated to cause <1% of ES [5]. The fused region of FLI1 or other ETS family transcription factors contains a conserved DNA-binding domain (DBD) responsible for the DNA binding of the fusion [2] and essential for the oncogenic

Received: July 18, 2024. **Revised:** February 13, 2025. **Editorial Decision:** March 1, 2025. **Accepted:** March 4, 2025

transcriptional program [6]. The structures of the DBDs of FLI1, ERG, FEV, ETV1, and ETV4 are similar, as defined by the X-ray crystallographic studies of these domains and their complexes with DNA from our laboratory and other groups [7–9]. In these proteins the DBD (residues 279–370 in FLI1; Supplementary Fig. S1) starts with helix α1 and ends with helix α4, which is not always present in other ETS family proteins [10]. The sequence homology of the DBDs is especially high for FLI, ERG, and FEV, which were previously grouped based on their DNA binding preferences [11]. In agreement with the structurally defined domain boundaries, deletion constructs of EWS–FLI1 fusions with a shortened C-terminal helix (α4) of the FLI1 DBD were impaired in transcriptional functions [12]. Divergent protein regions outside of the DBD in ETS family were found to be responsible for specific regulatory and autoregulatory functions in GABP, ETS1, ETS2, and ERG [8, 13–15].

ETS family transcription factors bind to the DNA sequence GGA(A/T). Depending on the sequences flanking the GGA(A/T) site, standalone GGA(A/T) sites can be high-affinity (such as CCGGAA; herein designated as "s" for "strong") or low-affinity (such as AAGGAA, designated as "w" for "weak") [16]. Some weak sites are located in proximity to binding sites of other transcription factors, such as Runt domain proteins, which can bind DNA cooperatively with FLI1, ERG [17], and ETS1 [11, 18, 19]. Other weak sites, instead, are found as contiguous repeats of GGAA sites, known as GGAA microsatellites, in promoters of some genes regulated by EWS–FLI1. An example of such a gene is *NR0B1*, whose upregulation by EWS–FLI1 is required for ES oncogenesis [20]. The EWS–FLI1 binding at the *NR0B1* promoter takes place on a GGAA microsatellite that consists of 25 contiguous GGAA sites; this binding is necessary and sufficient for the activation of the *NR0B1* gene [21]. GGAA microsatellites are overrepresented in promoters of genes directly bound by EWS–FLI1, where the enrichment is near the genes upregulated (but not downregulated) by EWS–FLI1 [21]. Multiple contiguous GGAA sites are necessary for the transcriptional activation by EWS–FLI1, and the degree of activation increases as the number of GGAA repeats increases. The size of an EWS–FLI1–DNA complex *in vitro* increases with the increasing number of contiguous GGAA sites in DNA oligomers of the same size, indicating that multiple copies of EWS–FLI1 are bound to these DNA ligands and, by inference, to GGAA microsatellites [21]. The stability of EWS–FLI1–GGAA microsatellite complexes was hypothesized to be a result of cooperative binding of multiple EWS–FLI1 proteins. These initial findings were corroborated and extended by other studies, revealing that EWS–FLI1 converted GGAA repeats into *cis*-regulatory elements acting as enhancers, but not repressors, by forming a nucleoprotein assembly of some sort that opened chromatin and recruited Pol II to drive the oncogenic transcriptional program [22–26 27]. Interfering with this assembly was deleterious to the oncogenic transcription, and it was suggested to be a potential therapeutic strategy [28].

Since the original discovery of the key oncogenic activity of EWS–FLI1 on GGAA microsatellites, its importance has been accentuated, but mechanistic information on this enigmatic assembly has remained elusive. Our study was designed to fill this gap and provide structural and functional details of the assembly of the FLI1 DBD on GGAA repeats.

Materials and methods

Protein preparation

Wild-type FLI1 (259–399) and FLI1^{F362A} (259–399) were cloned, expressed in *Escherichia coli* BL21 (DE3) (Agilent Technologies) and purified as we described previously [7]. The other FLI1 constructs used in this study were cloned and mutated analogously, using the primers given in Supplementary Table S1. The F362A mutation was introduced in all constructs, except FLI1 (276–371), which eluted from the S-200 size-exclusion column as a single peak corresponding to the monomeric protein. Additional single mutations N329E, D333G, and a double mutation N329E/D333G were introduced in FLI1^{F362A} (276–375) by the analogous procedure to that used previously for F362A. The same expression and purification protocols were then used. As in the previous study, the N-terminal decahistidine tag was removed by Prescission protease (GenScript) from all the proteins in the course of purification. In the last purification step, the proteins were purified on a Sephacryl S-200 size-exclusion column (GE Healthcare) in the gel filtration buffer (40 mM Tris–HCl, pH 8.0, 400 mM NaCl, and 2 mM β-mercaptoethanol). The proteins were concentrated in an Amicon Ultra-15 centrifugal filter unit with a 10,000 Da molecular weight cut-off (EMD Millipore) to 12 mg/ml, rapidly frozen in liquid nitrogen and stored at −80°C.

DNA preparation

The double-stranded DNA (dsDNA) oligonucleotides used in crystallization in complexes with FLI1 and in DNA binding assays were annealed using single stranded DNA oligonucleotides (Integrated DNA Technologies) shown in Supplementary Table S2. The oligonucleotides were dissolved in the annealing buffer (10 mM Tris, pH 8.0, 50 mM NaCl), and the concentrations were measured. Then each pair of complementary strands was annealed at 1:1 ratio as previously described [7].

Crystallization and crystal structure determination

For cocrystallization with the 2-site DNA, we used FLI1 (259–399) and FLI1^{F362A} (259–399); for co-crystallization with the 3-site DNA, we used FLI1^{F362A} (259–375), and for cocrystallization with the 4-site DNA, we used FLI1^{F362A} (259–399). FLI1 and the dsDNA (annealed as described above) were mixed at molar FLI1:DNA ratios of 2.1:1, 3.1:1, or 4.1:1 for 2-site, 3-site, and 4-site DNA, respectively, on ice and diluted with cold 20 mM Tris, pH 8.0 buffer to a final NaCl concentration <50 mM and incubated for 20 min on ice, followed by concentrating at 4°C to 5–8 mg/ml in a 0.5 ml Amicon centrifugal filter unit with a 3,000 Da molecular weight cut-off. The crystals were grown at 21°C by vapor diffusion in hanging drops set up by mixing of 1 μl of FLI1–DNA complex and 1 μl of the reservoir solution equilibrated against 1 ml of reservoir solution. The crystallization of the FLI1 DBD (residues 259–371) was carried out analogously. The crystals were then gradually transferred to cryoprotectant solutions, which were the same in composition as the reservoir solutions and additionally contained cryoprotectants, incubated for 20 min and then frozen in liquid nitrogen by rapid immersion. The reservoir solutions and cryoprotectants for each type of crystal are given in Supplementary Table S3.

The X-ray diffraction data for the crystals of FLI1–DNA complexes were collected at synchrotron beamline 22-ID at

the Advanced Photon Source, Argonne National Laboratory, at the X-ray wavelength of 1.0 Å. The data for the crystals of the FLI1 DBD were collected at beamline NYX at the National Synchrotron Light Source, Brookhaven National Laboratory, at the wavelength of 0.9794 Å. The data were indexed, integrated and scaled with HKL-2000 [29]. The data for the crystals of the 4:1 FLI1–DNA complex were severely anisotropic, and the weak data due to the anisotropy underwent ellipsoidal truncation by the online Diffraction Anisotropy Server [30]. The crystal structures of FLI1–DNA complexes were determined by molecular replacement with PHASER [31] by using as search models our previously reported, appropriately truncated crystal structure of a 1:1 FLI1 DBD–DNA complex (PDB ID: 5JVT) [32] in the first Phaser run and the FLI1 DBD from this structure in subsequent runs. For the crystal structure of the autoinhibited FLI1 DBA, the FLI1 DBD portion of the same published structure was used as a molecular search model. The crystal structures were then iteratively built with COOT [33] and refined with REFMAC [34]. The PDB accession codes, data collection and structure refinement statistics for the FLI1–DNA complexes and the FLI1 DBD are given in Supplementary Tables S4 and S5, respectively. Solvent accessible surface areas were calculated using program Surface Racer [35] with a solvent probe radius of 1.4 Å.

EMSA experiments

The 6-carboxyfluorescein (6-FAM)-labeled dsDNA was incubated at 500 nM with FLI1 DBD proteins at different FLI1:DNA stoichiometric ratios, as specified in the text, in the electrophoretic mobility shift assay (EMSA) buffer (20 mM Hepes, pH 7.8) on ice for 30 min. Samples with 5% glycerol were loaded on and run in a 6% nondenaturing polyacrylamide gel in 0.5× TBE running buffer (45 mM Tris, 45 mM boric acid, and 1 mM EDTA) at 4°C. The gels were scanned on a Typhoon FLA 9000 gel scanner (Cytiva). The EMSA experiments were performed at least in independent duplicates.

DNA-binding assays monitored by fluorescence anisotropy measurements

The FLI1 constructs and the DNA containing two weak sites as well as DNA containing a single strong and a single weak site used in the assays are described above and specified in Fig. 4 and Supplementary Table S2. FLI1 was titrated at specified concentrations into 6-FAM-labeled dsDNA at 20 nM in the DNA binding buffer (20 mM Hepes, pH 7.8, 40 mM NaCl, and 5 mM $MgCl_2$) and incubated at room temperature for 10 min. The experiments were performed at least in independent triplicate titrations. Fluorescence anisotropy was measured on a SpectraMax M5 microplate reader (Molecular Devices) with the excitation and emission wavelengths set at 494 and 525 nm, respectively. The data for the titrations using single-site DNA were analyzed using a 1:1 binding model, by a procedure reported previously [7], to obtain the best-fit curves and values for equilibrium dissociation constant K_d.

The data for the titrations using DNA with two weak sites were analyzed by using the 2:1 FLI1–DNA-binding equilibria:

$$\mathrm{D} + \mathrm{P} \overset{\mathrm{K_1}}{\rightleftarrows} \mathrm{DP} \tag{1}$$

$$\mathrm{DP} + \mathrm{P} \overset{\mathrm{K_2}}{\rightleftarrows} \mathrm{DP_2} \tag{2}$$

where DNA and FLI1 are denoted by D and P, respectively. The equilibrium association binding constants K_1 and K_2 are related to the microscopic site binding constant $K_{site} = 1/K_d$ by the following relationships [36]:

$$\mathrm{K_1} = 2\mathrm{K_{site}} \tag{3}$$

$$K_2 = (w\mathrm{K_{site}})/2 \tag{4}$$

where w is the cooperativity factor. Each protein binding event results in a relatively small fluorescent anisotropy change (far from reaching the theoretical maximum of 0.4); therefore, to a good approximation, the fluorescence anisotropy changes in binding events described by equilibria (1) and (2) can be assumed to be similar. Then, to a good approximation, fluorescence anisotropy (r) is related to total concentration of protein $[\mathrm{P}]_\mathrm{T}$ in the binding mixture by the following equation:

$$r = r_0 + a\left[2\mathrm{K_{site}}[\mathrm{P}]_\mathrm{T} + w\mathrm{K_{site}}^2([\mathrm{P}]_\mathrm{T})^2\right]/A \tag{5}$$

where r_0 is the anisotropy in the absence of protein, a is a scaling constant and

$$A = 1 + 2\mathrm{K_{site}}[\mathrm{P}]_\mathrm{T} + w\mathrm{K_{site}}^2([\mathrm{P}]_\mathrm{T})^2 \tag{6}$$

Equation (5) was used to carry out nonlinear regression fitting with SigmaPlot 9.0 (Systat) to obtain the best-fit curves and values of a and w, using the values of K_d from the analysis of the data collected with the single weak site DNA (Table 1). In the nonlinear regression, the average fluorescence anisotropy values were weighted by the respective reciprocal relative errors of the measurements.

Results

Crystal structures of the FLI1 DBD in complex with two contiguous GGAA sites

To unravel the architecture of FLI1–GGAA microsatellite complexes, we first focused on obtaining a crystal structure of the FLI1 DBD in complex with DNA containing two contiguous GGAA sites. We overexpressed and purified a FLI1 construct (residues 259–399) containing the FLI1 DBD and regions flanking it at the N- and C-termini. This construct is a part of oncogenic EWS–FLI1 fusions. We crystallized and determined a crystal structure of this protein in complex with a double-stranded 15-mer DNA 5′-GACCGGAAGGAAGTG-3′, at a resolution of 2.59 Å (Fig. 1A). We showed previously that the wild-type FLI1 DBD was in a slow equilibrium between a monomeric and a dimeric state [7]. The monomeric point mutant F362A, which was designed based on the crystal structure of the FLI1 dimer, was shown to be fully functional in DNA binding [7], transcriptional activation on microsatellites and oncogenic transformation (in the context of the EWS–FLI1 fusion) [12], strongly suggesting that the monomeric state is the functionally relevant state. For this reason, we also determined a crystal structure of this mutant in complex with the same DNA, at a resolution of 2.55 Å (Supplementary Fig. S2A). The wild-type and the monomeric mutant complexes with DNA crystallized in different crystal forms. In both crystals, there was one 2:1 protein–DNA complex per asymmetric unit involving end-to-end DNA crystal packing contacts and protein–protein contacts.

Both crystal structures revealed that two FLI1 DBDs were bound to the two GGAA sites of the DNA oligomer. Overall features of FLI1 DNA binding and site-specific recognition in these structures are similar to those in our previously reported structures of the 1:1 FLI1–DNA complex [7, 32];

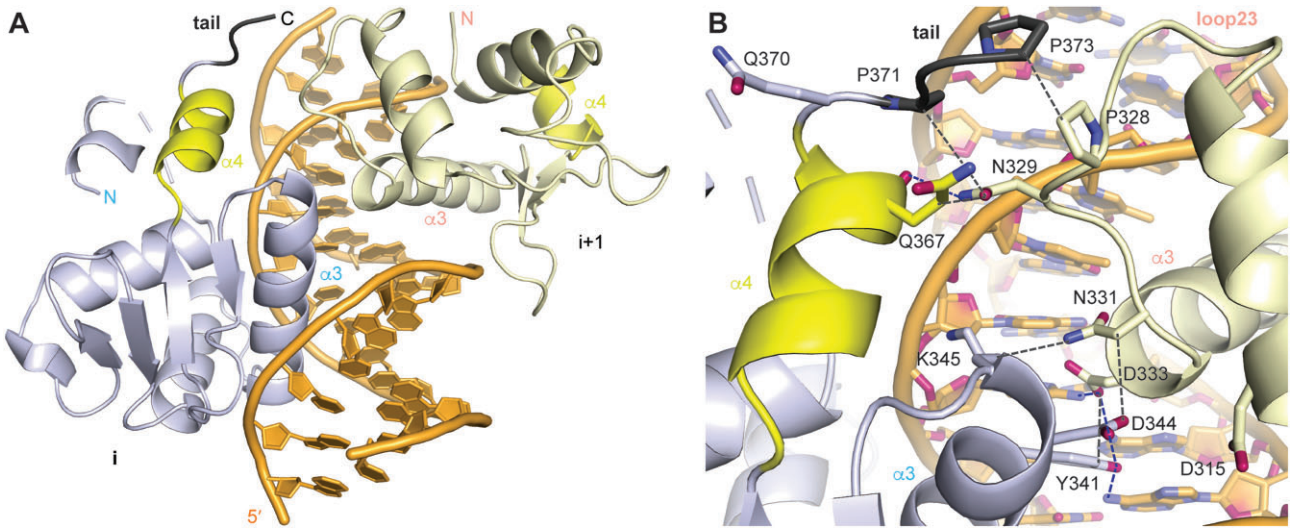


Figure 1. The crystal structure of the 2:1 FLI1 DBD–DNA complex on two contiguous GGAA sites. (**A**) The overall view of the complex. The C-terminal helix is colored yellow; the tail is black. (**B**) A magnified view of the interface between the two neighboring FLI1 DBDs. The interacting residues are shown as sticks and labeled. Hydrogen bonds are shown by dashed blue lines, steric and hydrophobic interactions are shown by dashed gray lines.

therefore, here we will describe only new findings related to binding of multiple FLI1s to contiguous GGAA sites. For the sake of clarity, we will refer to the FLI1 DBD bound to the first (from the 5′-end on the GGAA strand) and the second GGAA site as the *i*-th and the (*i* + 1)-th FLI1 (Fig. 1), respectively, and use subscripts *i* and *i* + 1 for their respective residues. In the structure of the FLI1^{F362A}–DNA complex, both FLI1 proteins packed against their crystallographic neighbors forming dimers as previously observed for wild-type FLI1 and ERG DBDs, when they were crystallized alone or in 1:1 complexes with one-site DNA [7, 8], despite the mutation (Supplementary Fig. S3A). In the structure of the wild-type FLI1 DBD with the two-site DNA, the (*i* + 1)-th FLI1 was also packed against its crystallographic symmetry mate in this dimeric form (Supplementary Fig. S3B). The region N-terminal to Ile281 was disordered in these dimeric FLI1 DBDs. In contrast, the *i*-th FLI1 of the 2:1 wild-type FLI1–DNA complex was packed against its symmetry mate in an unrelated arrangement, indicative of monomer–monomer crystal packing. In this interface, the N-terminal region (residues 268–274) folded as a short helix against the N-terminal helix of the DBD and made 2-fold symmetrical contacts with these regions of the symmetry mate, while residues 275–280 were disordered (Supplementary Fig. S3B). The different packings in the crystals demonstrated a dynamic nature of the monomer–dimer equilibrium of the wild-type FLI1, and they showed that even the mutant FLI1 DBD retained some propensity for dimerization at high protein concentrations used in the crystallographic experiments or only in the crystalline state. Therefore, the current structures indicated that both monomeric and dimeric forms of FLI1 could be accommodated in the 2:1 FLI1–DNA complex and that the dimeric state of FLI1^{F362A} was a crystallization artifact.

These structures of the 2:1 FLI1–DNA complex provide a proof of principle that the FLI1 DBD can occupy two contiguous GGAA sites simultaneously. In this ternary complex, the two major DNA recognition helices α3 of the two FLI1s form a 70° angle, with the end of the helix of the *i*-th FLI1 and the beginning of the respective helix of the (*i* + 1)-th FLI1 at the vertex of this angle, as expected based on the previous crystal structures of the 1:1 complex (Fig. 1A). Several residues in the DNA recognition helices of the two FLI1 proteins are in direct contact with each other in this complex, composing a large part of the FLI1–FLI1 interface. (Fig. 1B). One set of these interactions is between the side chains of Tyr341$_i$ and Asp333$_{i+1}$ in the protein–DNA interface, where one of the carboxyl group oxygens of Asp333$_{i+1}$ is engaged in weak bidentate hydrogen bonding with the hydroxyl group of Tyr341$_i$ and the N4 amino group of the C complementary to the first G of the second GGAA site. The hydroxyl group of Tyr341$_i$ is also involved in DNA recognition, forming a hydrogen bond with the N6 of the first A of the first GGAA site and a nonpolar contact with the T complementary to the second A. Additionally, the carboxyl group of Asp333$_{i+1}$ and the phenol ring of Tyr341$_i$ make nonpolar contacts with each other. Another set of contacts between the DNA recognition helices are steric contacts between Lys345$_i$ and Asp344$_i$ of the *i*-th FLI1 and Asn331$_{i+1}$ and Asp315$_{i+1}$ of the *i* + 1-th. The side chain of Asn331$_{i+1}$ interacts sterically with both Lys345$_i$ and Asp344$_i$. Asp315$_{i+1}$ interacts only with Asp344$_i$, both sterically and via a network of divalent metal ions and coordinated water molecules, some of which hydrate the phosphodiester backbone of the DNA. Lys345$_i$ is engaged in a salt bridge with the DNA backbone of the opposite strand. Other FLI1–FLI1 contacts differ between the wild-type and the mutant FLI1–DNA structures (Supplementary Fig. S2B). The last set of FLI1–FLI1 interactions in both structures involve the C-terminal region of the *i*-th FLI1 DBD contacting the (*i* + 1)-th FLI1. The conformation of the C-terminal region differs considerably between the wild-type (monomeric) and the mutant FLI1 (dimeric) structures. In the structure with the mutant FLI1, where the *i*-th FLI1 is in the dimeric state, its C-terminal helix α4 (residues 362–369) is engaged in swapped–helix interactions with the interacting protomer, as we previously reported [7]. In that structure, the side chains of Gln370$_i$ at the C-terminal end of the last helix of the *i*-th FLI1 and Asn329$_{i+1}$ in loop L23 are engaged in two reciprocal weak hydrogen bonds with one another. The region C-terminal to this helix (residues 371–399) is disordered. In the structure with the wild-type FLI1, where the *i*-th FLI1 is in a functional

monomeric state, the C-terminal helix of the *i*-th FLI1 is tilted by 20° away from its DBD and shifted along the helical axis toward the (*i* + 1)-th FLI1 by ~4 Å. Notably, a similar conformation of the C-terminal helix α4 was also observed in our recent crystal structures of the FLI1 DBD in complexes with Runx2 and DNA, and Runx2–Cbfβ and DNA [17], where FLI1 was also monomeric. This conformation of the helix could not be accommodated in the dimeric state. Residues $Gln367_i$ and $Ala368_i$ of the helix interact with $Asn329_{i+1}$ in loop L23 of the (*i* + 1)-th FLI1. Specifically, the C_α of $Ala368_i$ and the side chain of $Gln367_i$ interact sterically with the side chain of $Asn329_{i+1}$, and the backbone carbonyl oxygen of $Gln367_i$ is within a weak hydrogen bonding distance from the amino group of the carboxamide of $Asn329_{i+1}$. Furthermore, residues 371–373 C-terminal to this helix are ordered, and the backbone of this region forms a ~90° angle with the helical axis, so that $Pro371_i$ and $Pro373_i$ interact sterically and via nonpolar interactions with the side chains of $Asn329_{i+1}$ and $Pro328_{i+1}$ of the (*i* + 1)-th FLI1, respectively. We will call this region (residues 371–373) the C-terminal tail (even though there are other, disordered residues after residue 373), and this conformation of the tail the bracket conformation. All of the residues involved in the $FLI1_i$–$FLI1_{i+1}$ interactions are conserved in FLI1, ERG, and FEV (Supplementary Fig. S1), whose fusions are found in all but a few recently documented cases of Ewing sarcoma (containing ETV1 and ETV4). A few minor exceptions are noted in ETV1 and ETV4, which, for example, also contain a Pro analogous to Pro373, and an Ala in place of Pro371. The $FLI1_i$–$FLI1_{i+1}$ interactions on DNA bury 523 Å^2 of solvent accessible surface area (60% nonpolar and 40% polar) augmenting the solvent accessible area buried in the FLI1 monomer–DNA interface (1906 Å^2; 45% nonpolar, 55% polar) by 27%. The 60:40 nonpolar:polar buried solvent accessible area ratio is typical of protein–protein interactions. The increase in buried surface area from 1906 Å^2 by 523 Å^2 is consistent with a ~30-fold increase in binding affinity, as shown by the analysis of a large number of protein–protein interfaces with known affinities [37]. The nearest-neighbor interactions between the FLI1 proteins bound to adjacent GGAA sites are strongly suggestive of DNA-binding cooperativity.

Crystal structures of the FLI1 DBD in complex with three and four contiguous GGAA sites

Because functional GGAA microsatellites contain at least four GGAA sites [21], we sought to obtain structures of the FLI1 DBD on DNA oligomers containing more than two sites. In these experiments, we used the monomeric FLI1 mutant F362A, to ensure homogeneity of the complexes and to minimize dimer formation, which is not essential (and may not even be relevant) for the oncogenic transcriptional function of EWS–FLI1 *in vivo*. We crystallized and determined the crystal structures of the complexes of the FLI1 DBD with 19-mer and 25-mer DNA oligomers containing three and four contiguous GGAA sites, respectively (Fig. 2). It was increasingly challenging to obtain crystals of these FLI–DNA complexes as the number of GGAA sites increased, and to this end we had to test multiple DNA ligands with varying lengths and sequences of flanking DNA as well as various FLI1 DBD constructs. In both structures FLI1 occupied all available GGAA sites without gaps, forming nucleoprotein filaments. In describing these complexes, following the above nomenclature, we will denote by *i*, *i* + 1, *i* + 2, and *i* + 3 FLI1 proteins bound to consecutive sites counting from the 5′- to the 3′-end of the GGAA DNA strand.

Both structures contain one multi-protein–DNA complex per asymmetric unit, where the DNA molecules are not bent, consistent with previous EMSA studies [38]. In the structure of the 3:1 FLI1–DNA complex (FLI1, residues 259–275; Fig. 2A and B), each of the FLI1 proteins packs against another protein in a different asymmetric unit. The interactions of each protein with its GGAA site are similar to those observed in the 1:1 FLI1–DNA and in the 2:1 FLI1–DNA complexes. None of the three FLI1–FLI1 crystal packing arrangements resemble the previously observed and validated dimeric interface, indicating that all three FLI1 proteins are bound to DNA in the monomeric state. $FLI1_i$, $FLI1_{i+1}$, and $FLI1_{i+2}$ are all bound to DNA in the same conformation as that observed for $FLI1_i$ in the 2:1 wild-type FLI1–DNA structure described above, with the same tilt and shift of the C-terminal helix, and the C-terminal tail in the bracket conformation. Consequently, nearest-neighbor $FLI1_i$–$FLI1_{i+1}$, and $FLI1_{i+1}$–$FLI1_{i+2}$ interactions are similar to those described in detail above for the 2:1 wild-type FLI1–DNA structure (Fig. 1B). Additionally, steric interactions between $FLI1_i$ and $FLI1_{i+2}$ are also observed, with interatomic distances of 4.3–4.5 Å: side chains of $Gln280_i$ and $Ser326_i$ in loop L23 contact the side chain of $His352_{i+2}$ (Fig. 2C), which is additionally involved in electrostatic interactions with the DNA backbone.

The highest-order crystal structure that we obtained was the 4:1 FLI1–DNA complex (Fig. 2D). In this structure, DNA is packed in the crystal end to end, as in the previous structures, while $FLI1_{i+1}$ and $FLI1_{i+3}$ make crystal packing contacts with FLI1 proteins from another asymmetric unit that are also different from those in a FLI1 dimer. The other two FLI1s, $FLI1_i$, and $FLI1_{i+2}$, do not make crystal packing contacts. Therefore, all four FLI1 proteins are bound to the DNA in the monomeric state. Interestingly, despite this, $FLI1_i$ adopts a similar conformation of the C-terminal helix to that in dimeric FLI1 and does not contact $FLI1_{i+1}$. The tail region of $FLI1_i$ is disordered, apparently as a coupled conformational feature. Other interactions between $FLI1_i$ and $FLI1_{i+1}$ that are described above for nearest neighbors in other complexes are present. Even though no other significant conformational changes are present in either DNA or $FLI1_{i+1}$, a salient difference between the 4:1 and the 3:1 FLI1–DNA structures is the presence of $FLI1_{i+3}$ in the 4:1 FLI1–DNA complex. $FLI1_{i+3}$ interacts with $FLI1_{i+1}$ in this complex in the same manner as $FLI1_i$ interacts with $FLI1_{i+2}$ in the 3:1 complex, as described above. This interaction engages loop L23 of $FLI1_{i+1}$ of the 4:1 complex. The same loop would engage in interactions with the C-terminal helix and the tail of the adjacent upstream FLI1 neighbor in the 3:1 and 1:1 FLI1–DNA structures. Apparently, the $FLI1_{i+1}$–$FLI1_{i+3}$ interactions in the 4:1 complex replace the interactions between the C-terminal regions of $FLI1_i$ and $FLI1_{i+1}$ that are observed in the 3:1 complex. In FLI_{i+1} of the 4:1 complex, the C-terminal helix and the ordered tail in the bracket conformation interact with $FLI1_{i+2}$ as observed for the 3:1 complex. In $FLI1_{i+2}$, the C-terminal helix is in the conformation observed for the FLI1 dimer with the disordered tail; these regions do not interact with $FLI1_{i+3}$, while all other nearest-neighbor interactions are maintained. In this case, the observed conformation of the C-terminal region likely occurs due to steric hindrance by another complex in the crystal occupying the space needed for the interaction with $FLI1_{i+3}$.

Figure 2. The crystal structures of the 3:1 and the 4:1 FLI1–DNA complexes on three and four contiguous GGAA sites, respectively. (**A**) The overall view of the 3:1 FLI1–DNA complex. The C-terminal helix and the tail are shown in yellow and black. (**B**) An orthogonal view of the 3:1 complex. (**C**) A magnified view of the interactions between $FLI1_i$ and $FLI1_{i+2}$. The interacting residues are shown as sticks and labeled. Loop23 (L23) is also labeled. (**D**) The overall view of the 4:1 FLI1–DNA complex. Loop L23 of $FLI1_{i+1}$ engaged in the interaction with $FLI1_{i+3}$ is labeled.

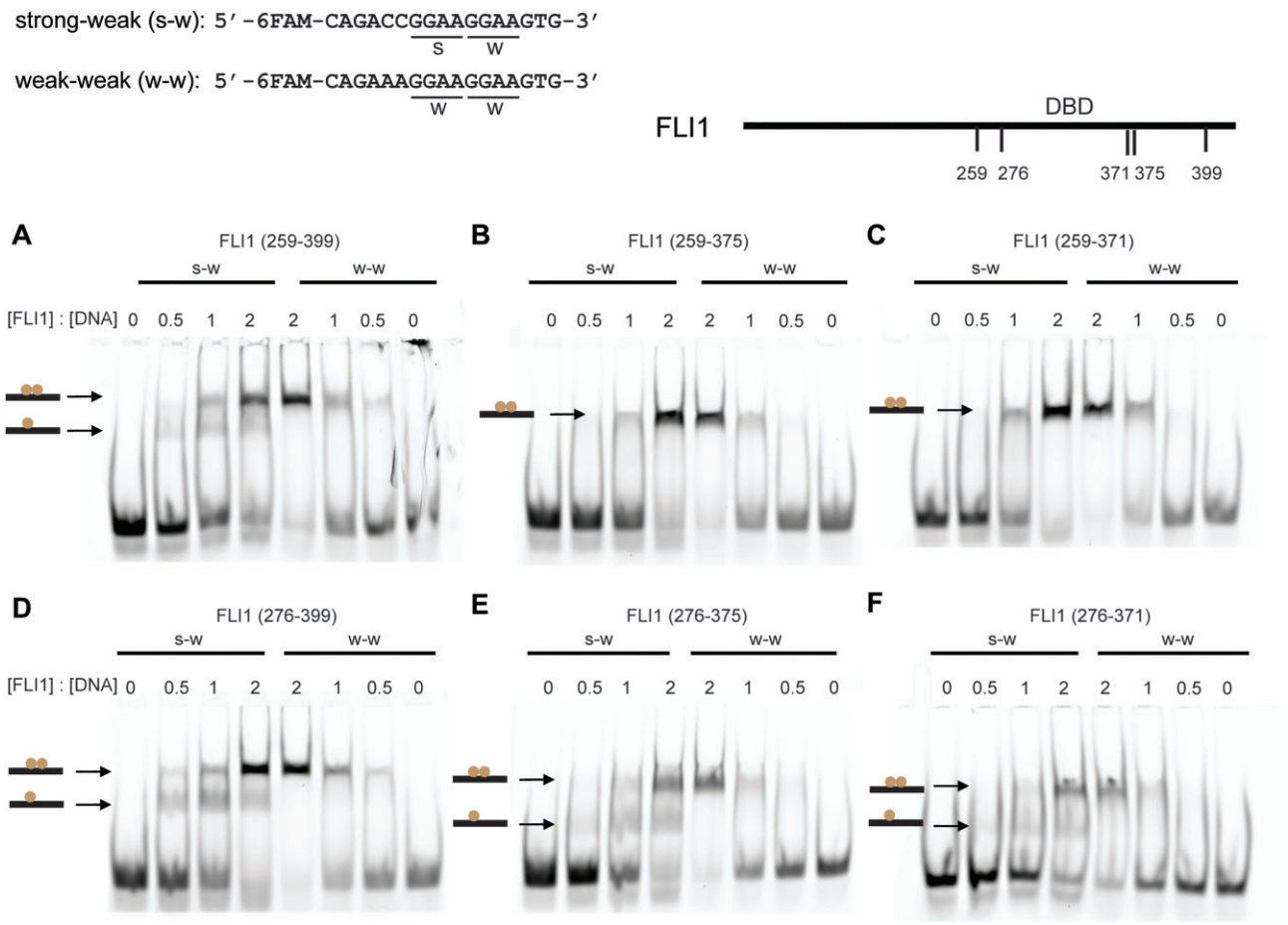


Figure 3. EMSA gel images of the titration of six FLI1 DBD constructs into two DNA oligomers, one containing contiguous strong (CCGGAA) and weak (AAGGAA) sites and one containing two contiguous weak sites (a microsatellite fragment). Panels (**A–F**) correspond to the six FLI1 constructs; the first four lanes in each panel contain the former DNA, the last four lanes contain the latter DNA. The DNA sequences and the FLI1 constructs endpoints are shown above the gel pictures.

Cooperative assembly of 2:1 FLI1–DNA complexes on contiguous GGAA sites observed by solutions studies

The structures of FLI1 on contiguous GGAA sites demonstrate that FLI1 proteins occupy all GGAA sites without gaps, and that in these complexes FLI1 interacts directly with its nearest neighbors and with second-nearest neighbors. These interactions are highly suggestive of DNA-binding cooperativity for FLI1 proteins in this assembly. To investigate if binding of multiple FLI1s to contiguous GGAA sites is cooperative and, if so, to determine the regions of FLI1 responsible for the cooperativity, we performed a series of EMSA experiments with fluorescently (6-FAM) labeled DNA oligomers bearing two contiguous GGAA sites (Fig. 3). We used two DNA oligomers: one, where the first site (on the 5′-end of the GGAA stretch) is a relatively low-affinity (weak; "w") binding site AAGGAA as is the other GGAA site, and the other, where the first site is a high-affinity (strong; "s") CCGGAA site. The second site is always weak (AAGGAA), since the 5′-AA is a part of the first GGAA site. We generated and used six constructs of the monomeric FLI1^{F362A} DBD, in which different N- and C-terminal flanking regions were deleted in different combinations (Fig. 3). Constructs starting at residues 259 and 276 were used to probe the effect of the N-terminal extension. For each of these N-termini, three constructs were generated, ending on residue 399, 375, and 371, to probe the effect of the truncation of the C-terminal extension to keep only the tail (up to residue 375) and to probe the effect of truncating the tail (ending on 371). These six FLI1 DBD constructs were tested with the two DNA sequences (w-w and s-w). A salient observation from this assay was that for the DNA with both weak sites (w-w) 1:1 FLI1–DNA complexes were not observed as FLI1 was titrated at constant DNA concentration (the right-hand side of the gel images). The 2:1 complex formed at [FLI1]:[DNA] ratios of 0.5 and higher for the constructs with the longest C-terminal flanking region (Fig. 3A and D) and at [FLI1]:[DNA] ratios of 1 and higher for the other constructs, which were truncated at the C-terminus, regardless of the presence of the N-terminal flanking sequence. This all-or-nothing behavior is direct evidence of cooperativity of binding of the FLI1 DBD to two individually weak contiguous GGAA sites. These observations also indicate that the DNA-binding cooperativity is compromised for the constructs truncated at the C-terminus. We could not discern the effect of a truncation of the tail (deleting residues 372–375) in this assay. The EMSA data with the DNA containing a high-affinity (strong) site (denoted as s-w; the left-hand side of the gel images) strongly corroborated these observations by showing the 1:1 complex that formed on the strong site for both constructs with the longest C-terminal flanking sequence

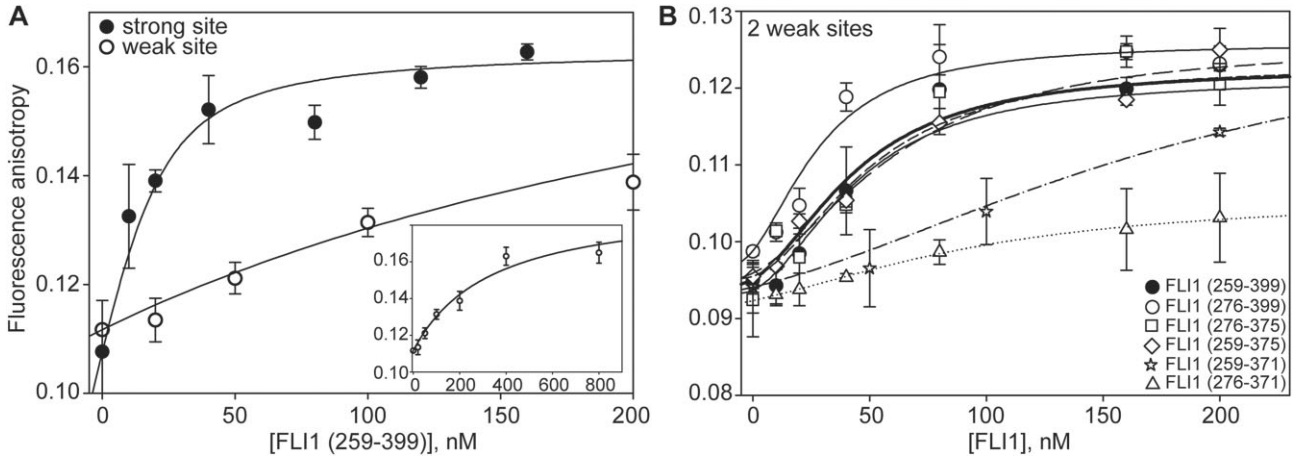


Figure 4. Equilibrium binding titrations of FLI1 into one-site and two-contiguous weak site DNA measured by monitoring fluorescence anisotropy of the probe on the DNA. (**A**) Titrations of the longest FLI1 DBD construct into one-site DNA (at 20 nM) containing a strong (filled circles) or a weak (open circles) site. (**B**) Titrations of the six FLI1 DBD constructs into DNA (at 20 nM) containing two contiguous weak sites. The curves correspond to the best fit to the 1:1 binding model (A) or the 2:1 cooperative binding model (B) with the parameter values given in the text and Table 1.

(ending on residue 399; Fig. 3A and D), and the 2:1 complex that dominated at the highest protein concentration, as expected. These data also revealed that the N-terminal flanking sequence (residues 259–275) was autoinhibitory, because truncating the C-terminal region weakened DNA binding of the constructs containing the N-terminal flanking sequence to the strong site (and likely to the weak site), so that the 1:1 complex was no longer observed (Fig. 3B and C), and the cooperative effect seen with the DNA containing two weak sites was exposed. In contrast, for the constructs with the truncated N-terminal region the 1:1 complex still formed, even when the C-terminal region was shortened (panels E and F).

To analyze quantitatively the cooperativity of FLI1 binding to DNA containing two contiguous weak GGAA sites, we measured fluorescence anisotropy using the same DNA with two weak sites as that used in the EMSA (defined as w–w in Fig. 3) upon titrating the six monomeric FLI1 DBD constructs (Fig. 4B). In order to interpret these data, we measured binding affinities (K_d) of the six monomeric FLI1 DBD constructs to a 6-FAM-labeled DNA oligomer containing a single weak (AAGGAA) site (Fig. 4A, Supplementary Fig. S4 and Table 1). As a control, we measured K_d for binding of the longest FLI1 DBD construct (residues 259–399) to DNA containing a single strong site (CCGGAA) site (Fig. 4A). The equilibrium binding constants for FLI1 (259–399) for a single strong (CCGGAA) and a single weak site obtained from these data were $K_d = 7.5 \pm 2.7$ nM and 320 ± 86 nM, respectively. The values of K_d for a single weak site displayed a 7-fold variation (137–1009 nM) among the FLI1 constructs. The N-terminal extension (259–275) was autoinhibitory, as also inferred from the EMSA analysis, with the constructs starting with residue 276 binding a single weak site with 2- to 3-fold increased affinity compared to their counterparts starting with residue 259. In contrast, truncating the C-terminal extension led to the loss of binding affinity; the majority of the affinity loss (2- to 3-fold) occurred upon truncating four residues (372–375) containing the C-terminal tail.

The titrations using the DNA with two weak sites (Fig. 4B) generally yielded apparent DNA-binding affinities that were ~10 times as strong as those to a single weak site for respective

Table 1. The DNA-binding affinity of the FLI1 DBD to a single weak GGAA site and the cooperativity factor for binding to two contiguous weak GGAA sites

FLI1 construct	K_d, single weak site	Cooperativity factor, w
259–399	320 ± 86	41 ± 21
259–375	544 ± 288	103 ± 26
259–371	1009 ± 421	22 ± 7
276–399	146 ± 37	16 ± 10
276–375	137 ± 34	33 ± 20
276–371	402 ± 172	11 ± 5
$276–375^{N329E}$	275 ± 39	6 ± 6
$276–375^{D333G}$	35 ± 4	~1
$276–375^{N329E/D333G}$	126 ± 14	~1

FLI1 constructs, indicative of strongly cooperative binding to the two DNA sites, in agreement with the EMSA data. Nonlinear regression analysis of these data using a 2:1 FLI1:DNA-binding model and the single-site affinities obtained above yielded the values of cooperativity factor (w; the fold increase in intrinsic DNA-binding affinity of FLI1 for a GGAA site when an adjacent site is occupied; Table 1) ranging from 10 to 103. These results indicate that the DBD lacking the N- and the C-terminal extensions (residues 276–371) is sufficient for cooperative binding. Two effects of the extensions could be seen from this analysis. First, the cooperativity was higher for constructs with the N-terminal extension than for those without one. Second, the C-terminal tail region (residues 372–375) contributed to the cooperativity regardless of the nature of the N-terminus. The latter effect is consistent with the nearest-neighbor interactions of the tail region observed in the crystal structures.

Probing the cooperative assembly of 2:1 FLI1–DNA complexes by point mutagenesis

To test the specific FLI1–FLI1 interfaces predicted by the structural studies to be involved in the DNA-binding cooperativity, we carried out point mutagenesis of these regions and measured DNA binding of the mutant FLI1 proteins by EMSA and fluorescence anisotropy measurements analogously to the as-

says described in the previous section, using the same two- and one-site DNA. We made two single (N329E and D333G) and one double (N329E/D333G) mutation in the FLI1 (276–375) construct. This construct exhibited both robust cooperativity and sufficiently high DNA-binding affinity to DNA containing a single weak GGAA site (Table 1), to allow for potential detuning of both properties by the mutations. We avoided mutating residues that made extensive interactions with DNA, to maintain DNA binding, which left us with Asn329 and Asp333. The N329E mutation in loop L23 enlarged this side chain. We predicted that the larger Glu on loop L23 would push the C-terminal helix α4 and the tail of the nearest neighbor on the DNA away (Fig. 1B), disrupting nearest-neighbor interactions between these structural elements. In addition, Glu329 would no longer provide a hydrogen bond donor for the main chain carbonyl of Gln367. The other mutation, D333G, was meant to perturb nearest-neighbor interactions between the DBD cores. Because the entire side chain of Asp333, including its C_β, is engaged in anion–π stacking onto the phenolic ring of the Tyr341 of the nearest neighbor (Fig. 1B), Asp333 had to be mutated to a Gly, removing these nonpolar interactions. Unavoidably, this mutation also removed a negative charge and eliminated a hydrogen bond with a DNA base, described previously in the text. All three mutant FLI1 constructs behaved similarly to the parent construct on the size-exclusion column during the last purification step, eluting at approximately the same time as a symmetrical peak at the expected volume for a monomer [7], with no significant aggregation and no self-association (Supplementary Fig. S5). This result indicated that the DBD fold was not disrupted by the mutations.

The EMSA demonstrated that the N329E mutation somewhat weakened the DNA-binding affinity to both strong and weak sites, while the DNA–binding cooperativity has persisted, as no 1:1 FLI1–DNA complex was observed for the DNA containing two weak sites (Fig. 5A). Fluorescence anisotropy measurements (Table 1 and Supplementary Figs S6 and S7) yielded a 5-fold smaller value of the cooperativity factor w ($w = 6 \pm 6$) for this mutant than for the respective wild-type construct ($w = 33 \pm 20$). This decreased cooperativity was similar to the cooperativity of the C-terminal tail truncation ($w = 11 \pm 5$), in agreement with predicted effect of this mutation in disrupting nearest-neighbor interactions involving the tail. The D333G mutation completely abolished the cooperativity, indicated by the appearance of the 1:1 FLI1–DNA complex with the w–w DNA (Fig. 5B). The fluorescence anisotropy measurements also indicated the loss of cooperativity, as w was indistinguishable from unity, while the binding affinity of this mutant to the DNA with a single site measured by this assay was ∼4-fold higher. The loss of cooperativity for the D33G mutant must have resulted from the disruption of core–core nearest-neighbor interactions, as predicted. This result, together with the data for the truncation mutants, indicate that the nearest-neighbor interactions between the FLI1 DBD cores are the dominant contributor to the DNA-binding cooperativity. In agreement with the single mutant data, the EMSA (Fig. 5C) and the fluorescence anisotropy data indicated that the double mutation abolished the DNA-binding cooperativity. The binding of the FLI1 double mutant to DNA containing a single weak site was similar to the respective wild-type FLI1 construct (Table 1), consistent with the opposite effects of the two single mutations.

The structural basis for autoinhibition by the N-terminal extension of the FLI1 DBD

The unbound FLI1 DBD that we structurally characterized previously [7] did not contain the N-terminal extension. Therefore, to establish the structural basis for the autoinhibition of DNA binding by the N-terminal extension, we determined a crystal structure of wild-type FLI1 (residues 259–371) containing this extension, at a resolution of 1.66 Å (Fig. 6A and Supplementary Table S5). Not surprisingly, this high-resolution crystal structure of this construct was very similar to the 2.1 Å resolution crystal structure of the autoinhibited DBD of ERG, nearly identical in sequence to the FLI1 DBD (PDB ID: 4IRH) [8]. Residues 272–280 were ordered, whereas residues 259–271 were disordered. The ordered region of the N-terminal extension of the FLI1 DBD was in an extended conformation interacting with the C-terminal helix of the DBD; a very similar conformation was observed for the ERG DBD. The superimposition of this FLI1 DBD structure and that of the *i*-th FLI1 DBD in the 2:1 wild-type FLI1–DNA complex (Fig. 6B) indicated that this conformation of the ordered portion of the extension directed the disordered part of the extension to clash sterically with bound DNA, explaining the autoinhibitory effect of the extension. Furthermore, the disordered part of N-terminal extension would also clash with loop L23 of the $i + 1$-th FLI1 DBD bound to the adjacent downstream GGAA site. To avoid these collisions in the above 2:1 FLI1–DNA complex, the N-terminal extension was rerouted away from the DNA and formed a short helix, whereas in the 2:1 $FLI1^{F362A}$ DBD–DNA complex, the region that would clash was directed away from the DNA and disordered (Supplementary Fig. S2).

Discussion

EWS–FLI1 is an abnormal DNA-binding factor that functions by several mechanisms, including gene repression and activation, each of which is necessary for the oncogenesis [39]. While the Ewing sarcoma tumorigenesis requires the FLI1 DBD, there are mechanisms involving the EWS region that do not directly involve the FLI1 DBD [40]. EWS–FLI1 assembly on GGAA microsatellites has been established as a mechanism essential for transcriptional reprogramming of several genes and for oncogenesis in Ewing sarcoma, but the molecular details of this assembly have remained elusive. The structures of 2:1, 3:1, and 4:1 complexes of the FLI1 DBD on contiguous GGAA sites and accompanying biochemical data presented here reveal the structural elements of this assembly and its stability. Specifically, these structure-function studies show that the FLI1 DBD proteins bind all GGAA sites without gaps, and this assembly is favored by the strong nearest-neighbor cooperativity of FLI1 binding observed for two contiguous GGAA sites. The biochemical data with the truncation mutants and point mutants of the FLI1 DBD in the nearest-neighbor FLI1–FLI1 interface corroborate the structural data in directly demonstrating the cooperativity and yield the quantitative contributions to the DNA-binding cooperativity of the individual protein regions: the DBD core, its N-terminal extension, the C-terminal helix, the C-terminal extension and the short C-terminal tail adjoining the core. Notably, the decrease in the DNA-binding cooperativity as a result of the N329E mutation, designed to disrupt the interaction of this residue with the C-terminal helix of the nearest neighbor, is

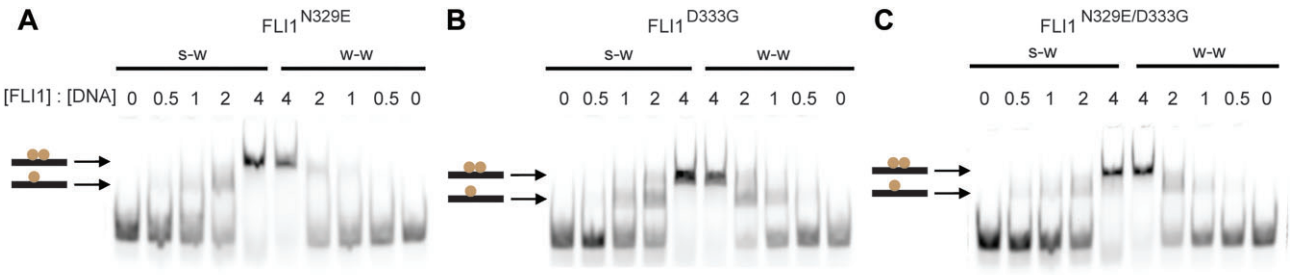


Figure 5. EMSA gel images of the titration of the mutant FLI1 (276–375) constructs into the two DNA oligomers described in Fig. 3. (**A**) The titration of $FLI1^{N329E}$. (**B**) The titration of $FLI1^{D333G}$. (**C**) The titration of $FLI1^{N329E/D333G}$.

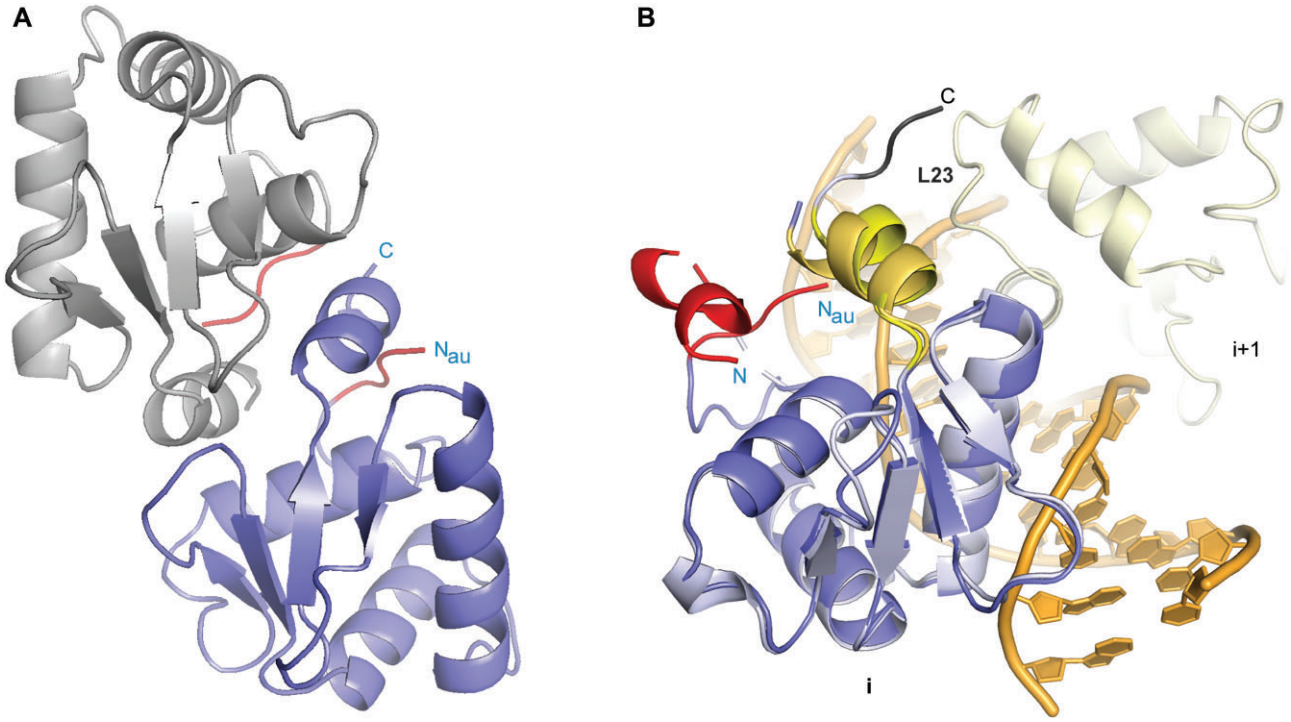


Figure 6. The structural basis for autoinhibition of DNA binding of the FLI1 DBD by its N-terminal extension. (**A**) The crystal structure of the dimer of the autoinhibited FLI1 DBD (residues 259–371). The C- and N-termini are denoted as C and N_{au} ("au" for autoinhibited), respectively. The N-terminal extension is in red. (**B**) The superimposition of the autoinhibited FLI1 DBD (as in panel A) onto the *i*-th FLI1 DBD (light blue) in the 2:1 FLI1–DNA crystal structure. The N-terminal extension of the *i*-th FLI1 DBD is partly helical (in red), with the N-terminus labeled as "N."

in agreement with recently reported transcriptional and DNA remodeling defects of EWS–FLI1 lacking this helix on GGAA microsatellites in Ewing sarcoma cells [41]. These effects of the C-terminal helix truncation were interpreted as a loss of cooperativity of EWS–FLI1 binding to GGAA microsatellites. Our structures may represent the intermediates in nucleation of the FLI1–DNA filament and its minimal growth. Because the FLI1 DBD shares high sequence and structural homology with the DBDs of the other ETS family transcription factors found in Ewing sarcoma fusions (ERG, FEV, ETV1, and ETV4; Supplementary Fig. S1), the structures and the mechanisms described in this study are likely common for this group of proteins.

It has been previously demonstrated that most of the GGAA repeat sequences contained four or more repeats; four being a minimum number required for the transcriptional activity of EWS–FLI1 [21]. Furthermore, the FLI1 DBD recapitulated the DNA binding and chromatin accessibility properties of the full-length EWS–FLI1 [12]. Therefore, our 4:1 FLI1–DNA structure serves as a mimic of a core of a minimally transcriptionally competent filament. Whether FLI1 occupies adjacent GGAA sites has been unclear prior to this study. Previous *in vitro* studies aimed to establish the protein:DNA-binding stoichiometry on GGAA repeats used a tagged EWS–FLI1 construct (Δ22), where all but the first six residues of the EWS portion were deleted, yielding conflicting results. DNaseI footprinting showed that with each GGAA added to a $(GGAA)_n$ sequence, the length of the protected region on the DNA increased by four bases, strongly indicative of occupancy at every site [38]. The same study showed that this construct did not bind two- and three-site DNA and bound with a 2:1 [protein]:[GGAA] stoichiometry DNA containing a larger number of sites, whereas a later study by the same group demonstrated binding to $(GGAA)_2$ and $(GGAA)_3$ DNA and yielded different (2:3, 1:2) [protein]:[GGAA] stoichiometries without cooperative binding, suggesting the presence of gaps [24]. It is

possible that the presence of the tag, the artificial sequence N-terminal to the FLI1 portion or the variable aggregation of this construct at the high concentrations used to determine DNA-binding stoichiometry have perturbed its assembly on DNA in these studies.

In agreement with the previous study of the ERG DBD [8], our DNA-binding assays reveal that the region N-terminal to the FLI1 DBD (residues 259–275) is autoinhibitory. The crystal structure of an unbound FLI1 DBD construct containing this region shows that this region is directed by its interactions with the C-terminal helix to clash with the DNA that would bind to the DBD core and with the FLI1 DNA bound to an adjacent downstream site. In the 2:1, 3:1 and 4:1 FLI1–DNA structures presented here, this region is disordered or folded into a short helix, away from the DNA. These structures suggest that the binding of the neighboring FLI1 may aid in rerouting the N-terminal extension away from the DNA. Then the seemingly surprising observation that the FLI1 constructs containing this N-terminal region display greater DNA-binding cooperativity than the constructs lacking this region could be explained by the relief of the autoinhibition upon binding DNA already occupied by other FLI1 molecules, due to some favorable, but dynamic interactions of this region with the neighboring FLI1. For ETS1, an ETS family transcription factor not implicated in Ewing sarcoma, strong autoinhibition by N- and C-terminal flanking regions is relieved by its interactions with itself, when two ETS1 DBDs are bound to a palindromic sequence containing two inversely oriented GGAA sites [42, 43], or by its interactions with other transcription factors bound to adjacent sites [19, 44–46].

The structures of 2:1, 3:1, and 4:1 FLI1–DNA complexes and the DNA binding data suggest the molecular logic and the directionality of FLI1–DNA filament assembly. These structures show that the interactions of the C-terminal tail of one FLI1 DBD (i-th) with loop L23 of the core DBD of the (i + 1)-th FLI1 bound immediately downstream (on the 3′-side) contribute to the strong FLI1–DNA-binding cooperativity, as shown by DNA-binding experiments for formation of 2:1 FLI1–DNA complexes. We propose that these intermolecular tail$_i$–core$_{i+1}$ interactions are significant for the nucleation of an EWS–FLI1–DNA filament, represented by the 2:1 and 3:1 FLI1–DNA complexes. As observed in the 4:1 FLI1–DNA complex, tail$_i$–core$_{i+1}$ interactions are absent if the (i + 1)-th FLI1 has the (i + 3)-th partner, because in this case loop L23 of the (i + 1)-th FLI1 is involved in the interactions with the core of the (i + 3)-th FLI1. These observations suggest a mechanism where FLI1–DNA filament grows on the 5′-end: starting with the 4-th FLI1, each newly bound FLI1 establishes interactions with its nearest neighbor and its second nearest neighbor without involving interactions of its tail or destabilizing other tail–core interactions. In this model, the tail–core interactions persist in the filament near its stationary 3′-end, stabilizing the three terminal FLI1 proteins. A mechanism where the filament propagates in both directions cannot be ruled out at this time.

What are the implications of this study for the assembly of the full-length EWS–FLI1 on microsatellites? Experiments in Ewing sarcoma cells published to date, strongly suggest that the binding of EWS–FLI1 to the GGAA microsatellites is cooperative [21, 41, 47], and that this cooperativity is mediated, at least in part, by the C-terminal helix [41], in agreement with our crystal structures and biochemical data. The FLI1 DBD is a feature present in both germline FLI1 and the EWS–FLI1 fusions, and it is necessary (as a part of the fusions), but not sufficient for the oncogenesis. The N-terminal EWS region in the fusions is also required for the oncogenic functions of EWS–FLI1 [48, 49]. This region contains low-complexity Gln- and Tyr-rich sequences that form intermolecular hydrogels or condensates [50]. These hydrogels were proposed to serve as a platform for binding of chromatin remodeling BRG1/BRM-associated factor (BAF) complexes [22]. Our structures of FLI1–DNA complexes, where the N-termini of FLI1 molecules point outward from the DNA, suggest that the low complexity EWS regions of at least four neighboring FLI1 DBDs could also assemble together by undergoing a coupled folding process to form a defined structure, rather than a completely disordered matrix. The EWS region may perturb the FLI1–DBD assembly or may simply augment it, allowing for regulation and dynamic interactions with binding partners and enabling the oncogenic functions [23]. The structural details of this larger assembly and its role in the transcriptional reprogramming are yet to be elucidated. We believe that this study will serve as a springboard to accessing high-order nucleoprotein complexes involving EWS–FLI1 and similar fusions.

Acknowledgements

The authors thank the staff of sector SER-CAT of the Advanced Photon Source at the Argonne National Laboratory and Dr. Kevin Battaile at beamline NYX at the Brookhaven National Laboratory for the assistance with the remote X-ray diffraction data collection. We also thank Dr. Jurgen Rohr, Dr. Markos Leggas, and Dr. Jon Thorson for insightful discussions.

Author contributions: C.H.: conceptualization, methodology, investigation, formal analysis, validation, writing- review and editing; O.V.T.: conceptualization, methodology, investigation, formal analysis, validation, project administration, funding acquisition, writing- original draft.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

The authors declare that they have no competing interests.

Funding

This work was funded by NIH grant R01CA243529. The synchrotron beamline access was supported by the Center for Structural Biology at the University of Kentucky. Funding to pay the Open Access publication charges for this article was provided by the Markey Cancer Center.

Data availability

The atomic coordinates and structure factor amplitudes were deposited into the Protein Data Bank (PDB). The PDB accession codes for each structure are specified in Supplementary Tables S4 and S5.

References

1. Van Mater D, Wagner L. Management of recurrent Ewing sarcoma: challenges and approaches. *OTT* 2019;**12**:2279–88. https://doi.org/10.2147/OTT.S170585
2. Delattre O, Zucman J, Plougastel B *et al.* Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature* 1992;**359**:162–5. https://doi.org/10.1038/359162a0
3. Crompton BD, Stewart C, Taylor-Weiner A *et al.* The genomic landscape of pediatric Ewing sarcoma. *Cancer Discov* 2014;**4**:1326–41. https://doi.org/10.1158/2159-8290.CD-13-1037
4. Sankar S, Lessnick SL. Promiscuous partnerships in Ewing's sarcoma. *Cancer Genet* 2011;**204**:351–65. https://doi.org/10.1016/j.cancergen.2011.07.008
5. Shing DC, McMullan DJ, Roberts P *et al.* FUS/ERG gene fusions in Ewing's tumors. *Cancer Res* 2003;**63**:4568–76.
6. May WA, Gishizky ML, Lessnick SL *et al.* Ewing sarcoma 11;22 translocation produces a chimeric transcription factor that requires the DNA-binding domain encoded by FLI1 for transformation. *Proc Natl Acad Sci USA* 1993;**90**:5752–6. https://doi.org/10.1073/pnas.90.12.5752
7. Hou C, Tsodikov OV. Structural basis for dimerization and DNA binding of transcription factor FLI1. *Biochemistry* 2015;**54**:7365–74. https://doi.org/10.1021/acs.biochem.5b01121
8. Regan MC, Horanyi PS, Pryor EE Jr *et al.* Structural and dynamic studies of the transcription factor ERG reveal DNA binding is allosterically autoinhibited. *Proc Natl Acad Sci USA* 2013;**110**:13374–9. https://doi.org/10.1073/pnas.1301726110
9. Cooper CD, Newman JA, Aitkenhead H *et al.* 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* 2015;**290**:13692–709. https://doi.org/10.1074/jbc.M115.646737
10. Hollenhorst PC, McIntosh LP, Graves BJ. Genomic and biochemical insights into the specificity of ETS transcription factors. *Annu Rev Biochem* 2011;**80**:437–71. https://doi.org/10.1146/annurev.biochem.79.081507.103945
11. Hollenhorst PC, Shah AA, Hopkins C *et al.* Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. *Genes Dev* 2007;**21**:1882–94. https://doi.org/10.1101/gad.1561707
12. Boone MA, Taslim C, Crow JC *et al.* The FLI portion of EWS/FLI contributes a transcriptional regulatory function that is distinct and separable from its DNA-binding function in Ewing sarcoma. *Oncogene* 2021;**40**:4759–69. https://doi.org/10.1038/s41388-021-01876-5
13. Fitzsimmons D, Lukin K, Lutz R *et al.* Highly cooperative recruitment of Ets-1 and release of autoinhibition by Pax5. *J Mol Biol* 2009;**392**:452–64. https://doi.org/10.1016/j.jmb.2009.07.028
14. Hagman J, Grosschedl R. An inhibitory carboxyl-terminal domain in Ets-1 and Ets-2 mediates differential binding of ETS family factors to promoter sequences of the mb-1 gene. *Proc Natl Acad Sci USA* 1992;**89**:8889–93. https://doi.org/10.1073/pnas.89.19.8889
15. Batchelor AH, Piper DE, de la Brousse FC *et al.* The structure of GABPalpha/beta: an ETS domain–ankyrin repeat heterodimer bound to DNA. *Science* 1998;**279**:1037–41. https://doi.org/10.1126/science.279.5353.1037
16. Wei GH, Badis G, Berger MF *et al.* Genome-wide analysis of ETS-family DNA-binding *in vitro* and *in vivo*. *EMBO J* 2010;**29**:2147–60. https://doi.org/10.1038/emboj.2010.106
17. Hou C, Mandal A, Rohr J *et al.* Allosteric interference in oncogenic FLI1 and ERG transactions by mithramycins. *Structure* 2021;**29**:404–12. https://doi.org/10.1016/j.str.2020.11.012
18. Shiina M, Hamada K, Inoue-Bungo T *et al.* A novel allosteric mechanism on protein-DNA interactions underlying the phosphorylation-dependent regulation of Ets1 target gene expressions. *J Mol Biol* 2015;**427**:1655–69. https://doi.org/10.1016/j.jmb.2014.07.020
19. Shrivastava T, Mino K, Babayeva ND *et al.* Structural basis of Ets1 activation by Runx1. *Leukemia* 2014;**28**:2040–8. https://doi.org/10.1038/leu.2014.111
20. Kinsey M, Smith R, Lessnick SL. NR0B1 is required for the oncogenic phenotype mediated by EWS/FLI in Ewing's sarcoma. *Mol Cancer Res* 2006;**4**:851–9. https://doi.org/10.1158/1541-7786.MCR-06-0090
21. Gangwal K, Sankar S, Hollenhorst PC *et al.* Microsatellites as EWS/FLI response elements in Ewing's sarcoma. *Proc Natl Acad Sci USA* 2008;**105**:10149–54. https://doi.org/10.1073/pnas.0801073105
22. Boulay G, Sandoval GJ, Riggi N *et al.* Cancer-specific retargeting of BAF complexes by a prion-like domain. *Cell* 2017;**171**:163–78. https://doi.org/10.1016/j.cell.2017.07.036
23. Chong S, Dugast-Darzacq C, Liu Z *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* 2018;**361**:eaar2555. https://doi.org/10.1126/science.aar2555
24. Johnson KM, Mahler NR, Saund RS *et al.* Role for the EWS domain of EWS/FLI in binding GGAA-microsatellites required for ewing sarcoma anchorage independent growth. *Proc Natl Acad Sci USA* 2017;**114**:9870–5. https://doi.org/10.1073/pnas.1701872114
25. Zuo L, Zhang G, Massett M *et al.* Loci-specific phase separation of FET fusion oncoproteins promotes gene transcription. *Nat Commun* 2021;**12**:1491. https://doi.org/10.1038/s41467-021-21690-7
26. Grunewald TG, Bernard V, Gilardi-Hebenstreit P *et al.* Chimeric EWSR1–FLI1 regulates the Ewing sarcoma susceptibility gene EGR2 via a GGAA microsatellite. *Nat Genet* 2015;**47**:1073–8. https://doi.org/10.1038/ng.3363
27. Riggi N, Knoechel B, Gillespie SM *et al.* EWS–FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. *Cancer Cell* 2014;**26**:668–81. https://doi.org/10.1016/j.ccell.2014.10.004
28. Chong S, Graham TGW, Dugast-Darzacq C *et al.* Tuning levels of low-complexity domain interactions to modulate endogenous oncogenic transcription. *Mol Cell* 2022;**82**:2084–97. https://doi.org/10.1016/j.molcel.2022.04.007
29. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997;**276**:307–26. https://doi.org/10.1016/S0076-6879(97)76066-X
30. Sawaya MR. Methods to refine macromolecular structures in cases of severe diffraction anisotropy. *Methods Mol Biol* 2014;**1091**:205–14. https://doi.org/10.1007/978-1-62703-691-7_15
31. McCoy AJ, Grosse-Kunstleve RW, Adams PD *et al.* Phaser crystallographic software. *J Appl Crystallogr* 2007;**40**:658–74. https://doi.org/10.1107/S0021889807021206
32. Hou C, Weidenbach S, Cano KE *et al.* Structures of mithramycin analogues bound to DNA and implications for targeting transcription factor FLI1. *Nucleic Acids Res* 2016;**44**:8990–9004. https://doi.org/10.1093/nar/gkw761
33. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004;**60**:2126–32. https://doi.org/10.1107/S0907444904019158
34. Murshudov GN, Skubak P, Lebedev AA *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 2011;**67**:355–67. https://doi.org/10.1107/S0907444911001314
35. Tsodikov OV, Record MT Jr, Sergeev YV. Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J Comput Chem* 2002;**23**:600–9. https://doi.org/10.1002/jcc.10061
36. Levandoski MM, Tsodikov OV, Frank DE *et al.* Cooperative and anticooperative effects in binding of the first and second plasmid Osym operators to a LacI tetramer: evidence for contributions of non-operator DNA binding by wrapping and looping. *J Mol Biol* 1996;**260**:697–717. https://doi.org/10.1006/jmbi.1996.0431

37. Chen J, Sawyer N, Regan L. Protein–protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci* 2013;**22**:510–5. https://doi.org/10.1002/pro.2230
38. Gangwal K, Close D, Enriquez CA *et al.* Emergent properties of EWS/FLI regulation via GGAA microsatellites in Ewing's sarcoma. *Genes Cancer* 2010;**1**:177–87. https://doi.org/10.1177/1947601910361495
39. Sankar S, Bell R, Stephens B *et al.* Mechanism and relevance of EWS/FLI-mediated transcriptional repression in Ewing sarcoma. *Oncogene* 2013;**32**:5089–100. https://doi.org/10.1038/onc.2012.525
40. Gorthi A, Romero JC, Loranc E *et al.* EWS–FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma. *Nature* 2018;**555**:387–91. https://doi.org/10.1038/nature25748
41. Bayanjargal A, Taslim C, Showpnil IA *et al.* DBD-alpha4 helix of EWSR1::FLI1 is required for GGAA microsatellite binding that underlies genome regulation in Ewing sarcoma. bioRxiv, https://doi.org/10.1101/2024.01.31.578127, 7 June 2024, preprint: not peer reviewed.
42. Lamber EP, Vanhille L, Textor LC *et al.* Regulation of the transcription factor ets-1 by DNA-mediated homo-dimerization. *EMBO J* 2008;**27**:2006–17. https://doi.org/10.1038/emboj.2008.117
43. Babayeva ND, Wilder PJ, Shiina M *et al.* Structural basis of Ets1 cooperative binding to palindromic sequences on stromelysin-1 promoter DNA. *Cell Cycle* 2010;**9**:3126–34. https://doi.org/10.4161/cc.9.14.12257
44. Goetz TL, Gu TL, Speck NA *et al.* Auto-inhibition of ets-1 is counteracted by DNA binding cooperativity with core-binding factor alpha2. *Mol Cell Biol* 2000;**20**:81–90. https://doi.org/10.1128/MCB.20.1.81-90.2000
45. Kim WY, Sieweke M, Ogawa E *et al.* Mutual activation of ets-1 and AML1 DNA binding by direct interaction of their autoinhibitory domains. *EMBO J* 1999;**18**:1609–20. https://doi.org/10.1093/emboj/18.6.1609
46. Gu TL, Goetz TL, Graves BJ *et al.* Auto-inhibition and partner proteins, core-binding factor beta (CBFbeta) and ets-1, modulate DNA binding by CBFalpha2 (AML1). *Mol Cell Biol* 2000;**20**:91–103. https://doi.org/10.1128/MCB.20.1.91-103.2000
47. Garcia-Aragoncillo E, Carrillo J, Lalli E *et al.* DAX1, a direct target of EWS/FLI1 oncoprotein, is a principal regulator of cell-cycle progression in Ewing's tumor cells. *Oncogene* 2008;**27**:6034–43. https://doi.org/10.1038/onc.2008.203
48. Lessnick SL, Braun BS, Denny CT *et al.* Multiple domains mediate transformation by the Ewing's sarcoma EWS/FLI-1 fusion gene. *Oncogene* 1995;**10**:423–31.
49. May WA, Lessnick SL, Braun BS *et al.* The Ewing's sarcoma EWS/FLI-1 fusion gene encodes a more potent transcriptional activator and is a more powerful transforming gene than FLI-1. *Mol Cell Biol* 1993;**13**:7393–8.
50. Johnson CN, Sojitra KA, Sohn EJ *et al.* Insights into molecular diversity within the FUS/EWS/TAF15 protein family: unraveling phase separation of the N-terminal low-complexity domain from RNA-binding protein EWS. *J Am Chem Soc* 2024;**146**:8071–85. https://doi.org/10.1021/jacs.3c12034

Received: July 18, 2024. **Revised:** February 13, 2025. **Editorial Decision:** March 1, 2025. **Accepted:** March 4, 2025