THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 271, No. 38, Issue of September 20, pp. 23329–23337, 1996

Printed in U.S.A.

New Insights on DNA Recognition by ets Proteins from the Crystal Structure of the PU.1 ETS Domain-DNA Complex*

(Received for publication, May 24, 1996, and in revised form, July 6, 1996)

Frédéric Pio‡§, Ramadurgam Kodandapani‡§, Chao-Zhou Ni‡, William Shepard¶, Michael Klemsz‖, Scott R. McKercher‡, Richard A. Maki‡, and Kathryn R. Ely‡ ‡‡**

*From the ‡La Jolla Cancer Research Center, The Burnham Institute, La Jolla, California 92037, ¶LURE, Université Paris-Sud, Bât. 209, 91405 Orsay Cedex, France, and **Neurocrine Biosciences, San Diego, California 92121*

Transcription factors belonging to the ets family regulate gene expression and share a conserved ETS DNA-binding domain that binds to the core sequence 5′-(C/A)GGA(A/T)-3′. The domain is similar to $\alpha+\beta$ ("winged") helix-turn-helix DNA-binding proteins. The crystal structure of the PU.1 ETS domain complexed to a 16-base pair oligonucleotide revealed a pattern for DNA recognition from a novel loop-helix-loop architecture (Kodandapani, R., Pio, F., Ni. C.-Z., Piccialli, G., Klemsz, M., McKercher, S., Maki, R. A., and Ely, K. R. (1996) *Nature* 380, 456–460). Correlation of this model with mutational analyses and chemical shift data on other ets proteins confirms this complex as a paradigm for ets DNA recognition. The second helix in the helix-turn-helix motif lies deep in the major groove with specific contacts with bases in both strands in the core sequence made by conserved residues in α3. On either side of this helix, two loops contact the phosphate backbone. The DNA is bent (8°) but uniformly curved without distinct kinks. ETS domains bind DNA as a monomer yet make extensive DNA contacts over 30 Å. DNA bending likely results from phosphate neutralization of the phosphate backbone in the minor groove by both loops in the loop-helix-loop motif. Contacts from these loops stabilize DNA bending and may mediate specific base interactions by inducing a bend toward the protein.

Transcription factors bind to target DNA sequences to regulate metabolic functions such as growth and differentiation. Typically, the molecular scaffold for DNA recognition is conserved within a given family of DNA-binding proteins. In some cases the similarity of these scaffolds suggests an evolutionary relationship between different families or comparison of scaffolds reveals a structural similarity that was obscured by sequence comparisons alone.

A recently discovered family of regulatory proteins, the *ets* gene family, includes more than 45 members in a variety of organisms from *Drosophila* to humans (1, 2). These molecules play a role in normal development and have been implicated in malignant processes such as erythroid leukemia and Ewing's sarcoma. The DNA-binding domain of ets proteins is a conserved region (ETS domain) that is about 85 residues in length. Although ets proteins share a homologous sequence in the ETS domain, they differ in length and in the relative position of this domain. In some molecules, the ETS domain is found at the carboxyl terminus (*e.g.* PU.1 (3); ets-1 (4); ets-2 (5)), while in others the domain is located in the middle of the sequence (erg (6)), or in the amino-terminal region (elk-1 (7)). Flanking regions are thought to form other functional domains that influence protein-protein recognition or inhibitory domains that mask the DNA-binding site (8, 9).[1] In ets-1, an α-helix that is located in an inhibitory domain immediately NH_2-terminal to the ETS domain unfolds on DNA-binding (10). Regardless of the position of the ETS domain within the intact ets proteins, there is strong sequence homology in this conserved region.

We have determined the crystal structure of the ETS domain of the PU.1 transcription factor complexed to DNA (11). The domain is similar to $\alpha+\beta$ helix-turn-helix (HTH)[2] DNA-binding proteins and contacts a 10-base pair region of duplex DNA that is bent (8°) but uniformly curved without distinct kinks. The PU.1 domain assumes a tight globular structure with three α-helices and a four-stranded antiparallel β-sheet enclosing a hydrophobic core. The topology of the domain is similar to the structures of other ets family proteins fli-1 (12), murine ets-1 (13), and human ets-1 (14) determined in solution by NMR. The common molecular scaffold is similar to DNA-binding proteins such as CAP (15) and resembles "winged"-HTH proteins including HNF-3γ (16). ETS domains bind as a monomer to the core sequence 5′-(C/A)GGA(A/T)-3′.

The PU.1 domain contacts DNA from three sites: the recognition helix (α3) interacts with the GGAA core sequence in the major groove, while contacts with the phosphate backbone on either side of this site are made in the minor groove by two loops. Therefore, the PU.1 ETS domain binds DNA by a loop-helix-loop motif. One loop is formed between β-strands 3 and 4 (a "wing") and the other is a loop in the position of the turn in the HTH motif (α2-turn-α3). The protein-DNA contacts stabilize a uniform bending of the duplex DNA that likely is due to phosphate neutralization by the PU.1 domain. Surprisingly, the protein-DNA interactions reported in the NMR structure of a human ets-1-DNA complex (14) differed dramatically from this pattern, involving different contacts and significant DNA deformation. Because of this discrepancy, we chose to test the validity of the PU.1-DNA complex as a model for other ets proteins. As reported here, when the results of mutational analyses on a number of ets proteins are correlated with the structure of the PU.1-DNA complex and with chemical shift data measured with the fli-1 (12) and murine ets-1 (13) mole-

* This work was supported in part by USAMRDC Department of the Army Grants DMD17–94J-4439 (to K. R. E.), National Institute of Allergy and Infectious Disease Grant AI20194 (to R. A. M.), and National Cancer Institute Grant CA63489 (to K. R. E.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Contributed equally to the results presented in this study.

‖ Present address: Dept. of Microbiology and Immunology, Indiana University School of Medicine, Indianapolis, IN 46202.

‡‡ To whom correspondence should be addressed: The Burnham Institute, 10901 N. Torrey Pines Rd., La Jolla, CA 92037. Tel.: 619-646-3135; Fax: 619-646-3196.

[1] M. Klemsz and R. A. Maki, unpublished results.

[2] The abbreviation used is: HTH, helix-turn-helix.

TABLE I
Crystallographic refinement statistics

R_{sym} (%)	3.3
Resolution range (Å)	6–2.1
Average B (Å^2)	31.65
Crystallographic R-factor (%)	22.5
R_{Free} (%)	28.7
Number of reflections used	22022 F $>2\sigma$(F)
Number of protein atoms	1486
Number of DNA atoms	1300
Number of solvent atoms	143

Root mean square deviation from ideal	r.m.s.	Target
Bond distance (Å)	0.012	(0.06)
Bond angles (degrees)	1.629	(10)
Dihedral angles (degrees)	1.575	(20)

cules, the loop-helix-loop scaffold is confirmed as a general model for DNA recognition by ets proteins. This pattern defines a new class of HTH DNA-binding proteins. The molecular pattern of DNA recognition by ets proteins is compared to other HTH proteins for which crystal structures of the protein-DNA complexes are available.

EXPERIMENTAL PROCEDURES

PU.1 DNA Complex—A recombinant fragment encompassing residues 160–272 from the murine ets protein PU.1 was crystallized in complex with a 16-base pair oligonucleotide representing a consensus PU.1 DNA-binding site (3) as described previously (17). The complex crystallized in space group $C2$ with $a = 89.1$, $b = 101.9$, $c = 55.6$ Å, and $\beta = 111.2°$. There are two complexes in the asymmetric unit. The length of the oligonucleotide was critical for crystallization and the oligonucleotide used to form the complex permitted end-to-end stacking of the DNA in the crystal lattice with the formation of pseudo-base pairing by the overhanging A and T bases.

Crystallographic Analyses—The initial structure analysis of the complex solved by the MIRAS method was reported (11). For this first phase of the study, a native data set and four heavy atom data sets were collected using a Rigaku RU200 rotating anode x-ray source and two San Diego Multiwire Systems area detectors. The initial data sets were collected from flash frozen crystals at 2.3-Å resolution. To refine the structure further, another native data set extending to 2.1 Å was collected at the LURE synchrotron source in Orsay, France. Diffraction data were collected at station D41 interfaced with the Mark III multiwire proportional area detector. Data sets were processed using MOSFLM (18) and ROTAVATA, AGROVATA, and TRUNCATE in the CCP4 package (19). In the present study, this native data set was scaled to the data collected in the home laboratory by Wilson scaling and the synchrotron data were incorporated into the refinement. The programs PHASES (20), FRODO (21), and X-PLOR (22) were used for structure solution, model building, and refinement. The current R-factor is 22.5 for 6 to 2.1 Å data (22,022 reflections). The average overall B-factor for 2929 non-hydrogen atoms (1486 protein atoms + 1300 DNA atoms + 143 solvent oxygens) is 31.6 Å^2. The refinement statistics are presented in Table I. There were 11 disordered residues at the amino terminus of the domain and 14 disordered residues at the carboxyl terminus of the recombinant fragment that were excluded from the model. These residues were not ordered even when the resolution was extended to 2.1 Å. For all residues representing the complete ETS domain (residues 171–258), the electron density was clear and permitted unambiguous fitting of both backbone and side chain atoms. More solvent atoms have been added to the model. Only minimal changes in the configuration of some side chains were evident in the high resolution map. The stereochemistry of all main chain torsion angles in the domain fall within energetically favorable limits (Fig. 1) indicating that no segment of the domain is denatured or randomly configured. The DNA was clearly defined even in the first MIRAS map.

Analyses of DNA Helical Parameters—To analyze the stereochemical basis for the uniform bending observed in the oligonucleotide bound in complex to PU.1, the DNA superstructure was measured (23, 24) and four parameters were calculated that describe the conformation of the DNA bases and the phosphate backbone. The values were calculated (excluding the 5′ A overhang) to analyze helical parameters along the length of the oligonucleotide and to compare these with standard B-DNA parameters. The geometry of dinucleotide steps was analyzed for three rotational angles defining twist, tilt, or roll and for one translational distance, *i.e.* rise. The values for these parameters are presented in Table II.

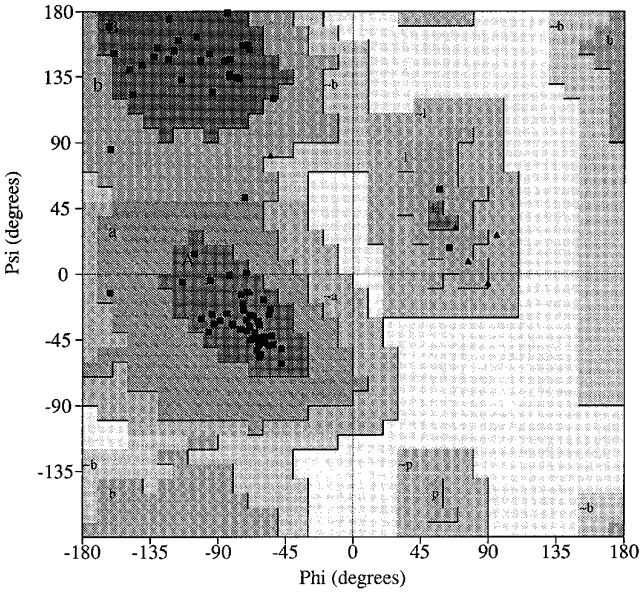

FIG. 1. **Ramachandran diagram for the current model of the PU.1 ETS domain.** This diagram presenting $\phi\psi$ angles (46) was produced using the PROCHECK programs (47). Glycine residues are represented by *triangles*. Various regions of the plot with different levels of shading are indicated with the *darkest shaded areas* corresponding to the energetically most favorable $\phi\psi$ angles.

Sequence Alignments and Structural Comparisons—Sequence alignments for ets proteins were made using GENEWORKS. The individual sequences were collected from the SWISSPROT data base and regions corresponding to the ETS domains were excised from the full-length protein before the alignment process began (25). The results of this comparison are presented in Fig. 2. Sequence comparisons between members of different families of HTH proteins were made using the program QUANTA (Molecular Simulations, Inc.) especially when structure-based alignments were utilized. To search structure data bases to identify proteins with similar overall scaffolds to the PU.1 domain, the algorithm DALI developed by Chris Sander (26) was used. For structural comparisons of HTH proteins, coordinates were obtained from the Brookhaven Protein Data Bank (27): 434 cro repressor (code 3CRO), λ repressor (code 1LMB), CAP (code 1CGP), and heat shock factor (code 2HTS). The coordinates for HNF-3γ were kindly provided by Dr. S. Burley. The actual structural comparisons/graphical analyses were performed using Quanta (Molecular Simulations, Inc.) and the Alberta/Caltech program TOM based on FRODO (21).

RESULTS AND DISCUSSION

The similarity of the structural organization of the ETS domains of PU.1 (11), fli-1 (12), and ets-1 (13, 14) and the presence of a conserved hydrophobic core suggests that this overall scaffold will be highly conserved in all members of the family. To facilitate comparisons, the sequences of the ETS domains of 33 members of the ets family are aligned (Fig. 2). The sequences of this domain in a number of ets proteins are identical for two or more species, representing a significant level of homology within the family. The results of mutational substitutions in a number of ets proteins are tabulated in Table III.

Hydrophobic Core—The importance of the hydrophobic core was verified by site-directed mutagenesis of the PU.1 domain (11). Of the 14 strictly conserved residues in the domain, seven are found in the hydrophobic core. Single substitution of glycine for five of these residues in PU.1 (Fig. 3) resulted in loss of DNA binding. Two of these core residues also contact the DNA phosphate backbone. The peptide amide nitrogen of Leu^{174} interacts with O2P of C-22 and the side chain NE-1 of Trp^{215} forms a hydrogen bond with O1P from T-23. Mutation of tryp-

TABLE II

DNA helical parameters of the 16-base pair oligonucleotide bound to PU.1

DNA structural parameters were refined in X-PLOR (22) and then analyzed using the programs developed by Babcock and Olson (24). For comparison, typical twist angles for B-DNA are 34.3°, roll angles are 0°, and rise values are 3.38 Å.

	Base pair	Inter-base pair			Slide (Å)	Intra-base pair	
		Helical twist (°)	Roll (°)	Rise (Å)		Propeller twist (°)	Buckle (°)
1	A-T	36.09	−0.06	3.18	0.13	−18.22	11.26
2	A-T	39.67	−0.48	3.23	0.06	−16.18	13.88
3	A-T	34.10	−6.19	3.30	−0.71	−14.38	1.39
4	A-T	36.09	0.76	3.20	−1.00	−17.25	3.38
5	G-C	27.50	6.59	3.57	−0.75	−8.02	2.87
6	G-C	29.06	7.29	3.11	0.24	2.10	−17.28
7	G-C	33.02	3.96	3.47	0.59	7.77	5.90
8	G-C	27.44	6.75	3.21	−0.37	−14.98	7.43
9	A-T	36.02	9.00	3.10	0.16	−21.87	10.59
10	A-T	39.21	3.24	3.35	−0.61	−19.29	5.58
11	G-C	24.41	−0.84	3.38	−0.94	−13.13	−10.09
12	T-A	37.07	3.75	3.29	0.98	−12.30	−7.49
13	G-C	32.70	9.93	3.30	−0.14	−6.40	3.66
14	G-C	33.29	4.99	3.23	−0.02	−10.53	−8.09
15	G-C					−9.96	−2.26

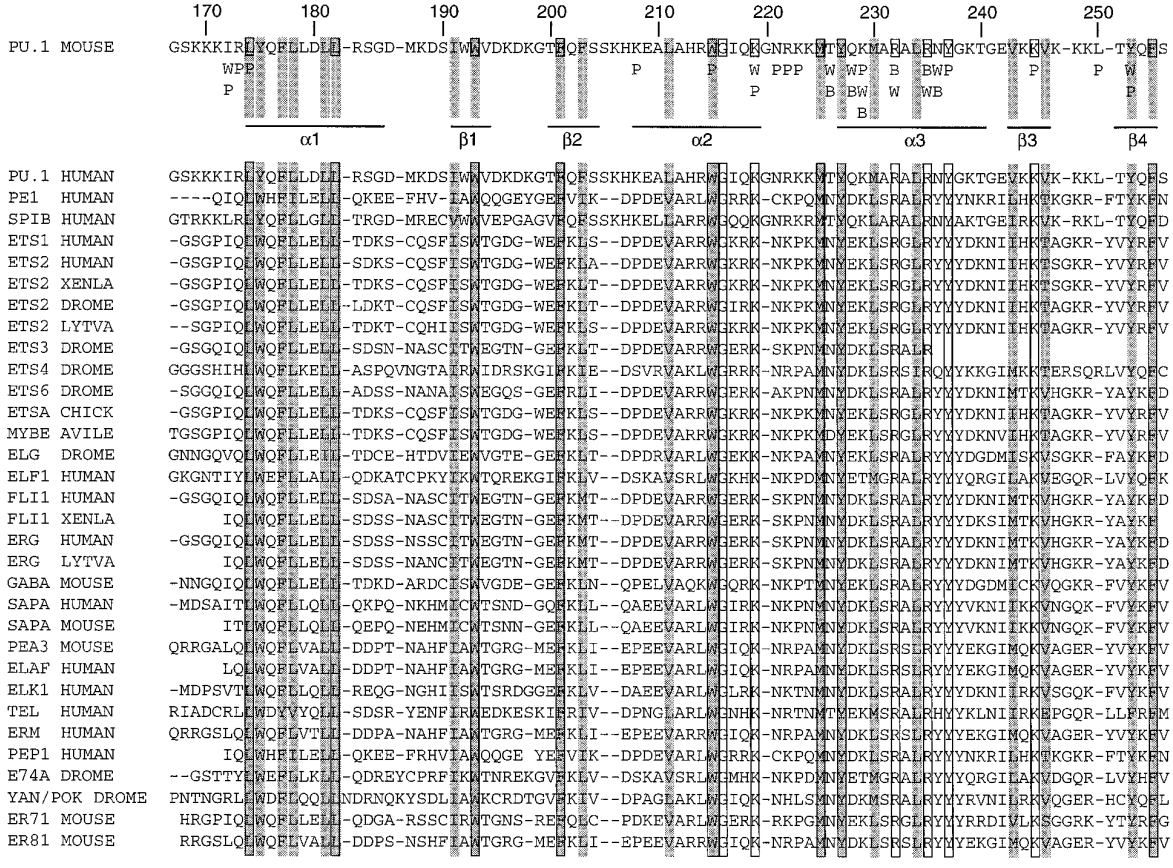

FIG. 2. **Sequence alignment of the DNA-binding domain of 33 members of the ets family.** The amino acid sequence of PU.1 is listed at the top of the figure and residues that are strictly conserved in the family are enclosed in *boxes*. The sequences were obtained from the SWISSPROT data base and original citations for the sequences are given in the data base. Secondary structural features of the PU.1 ETS domain are indicated above the alignment. Directly under the PU.1 sequence, the residues that contact DNA are indicated: *B*, base interaction; *P*, phosphate backbone interaction; *W*, water-mediated interaction. Residues found in the hydrophobic core in PU.1 and expected to be located in the hydrophobic interior of all ets proteins are *shaded*. In some cases, the sequences for ets proteins for two or several species are identical, and therefore only one sequence has been listed to avoid duplication.

tophan 215 to arginine results in loss of DNA binding in ets-1 (28, 29; see Table III). Substitutions in the hydrophobic core affect DNA binding probably because the changes disrupt the tight globular structure of the domain. Residues 174 and 215 are doubly critical for DNA binding since they represent both important structural residues in the domain core and actual DNA contact residues. In summary, residues in the hydrophobic core are critical for the formation of the overall scaffold for ets recognition.

Molecular Scaffold of ETS Domains—To evaluate the conservation of this scaffold within the ets family, the α-carbon backbones of PU.1 (11) and fli-1 (12) domains were superimposed utilizing both sequence homology and secondary structure similarities. For this purpose, a single model from the ensemble of structures deposited in the data bank was used for the NMR-derived fli-1 structure. This scaffold provides the

TABLE III
Mutations in the DNA-binding domain of ets family proteins that abolish DNA binding
The reference for each mutational substitution is given in parentheses with the protein studied.

ets Protein	Residue in PU.1[a]	Single mutation	Multiple mutations
PU.1 (11)	174H, D	L → G	
PU.1 (11)	178H	L → G	
ets-1 (29)	Multiple		174H, D, 175H, 177H, 178H
ets-1 (28)	185	K → P	
ets-1 (28)	191H	I → T	
PU.1 (11)	193H	W → G	
ets-1 (28)	194	T → I	
ets-1 (28)	196	D → G	
ets-1 (28)	201H	F → L	
PU.1 (11)	201H	F → G	
PU.1 (11)	203H	F → G	
ets-1 (28)	212	A → V	
ets-1 (28)	214	R → G	
ets1 (28, 29)	215H, D	W → R	
PU.1 (11)	215H, D	W → G	
ets-1 (28)	219D	K → *X*[b]	
PU.1 (11)	219D	K → G	
ets-1 (28)	222D	K → *X*[b]	
ets-1 (28)	227H	Y → C	
fli-1 (12)	228D	D → H, Q, K	
ets-1 (28)	232D	R → *X*[b]	
fli-1 (12)	232D	R → D, K, N	
PU.1 (11)	232D	R → G	
fli-1 (29)	234H	L → V	
ets-1 (29)	Multiple		234H, 235, 236, 237
fli-1 (12)	235D	R →K, D, N, E	
PU.1 (11)	235D	R → G	
fli-1 (12)	236D	Y → V	
ets-1 (48)	242	I → E, G, P, V	
ets-1 (28)	243H	I → T	
PU.1 (11)	245D	K → G	
ets-1 (28)	248	K → I	
ets-1 (28)	254H	F → L	

[a] Residue numbers of the PU.1 sequence are given to facilitate direct comparison with the sequence alignment in Fig. 2; H indicates a residue in the hydrophobic core of the PU.1 domain and D indicates residues which contact DNA in the PU.1-DNA complex, either directly or by water-mediated interactions.
[b] *X*, substitution by any amino acid.

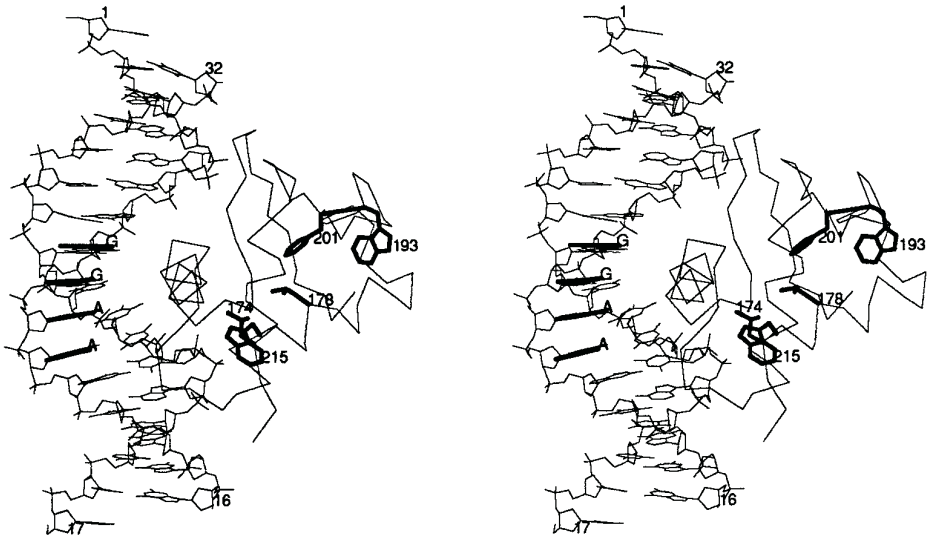

FIG. 3. **Stereodiagram of the PU.1-ETS domain DNA complex.** The α-carbon backbone for residues 171–258 is shown bound to DNA with the bases in the GGAA core in *bold lines*. The ETS module is composed of three α-helices and a four-stranded antiparallel β-sheet enclosing a hydrophobic core. There are seven strictly conserved residues in this core (Fig. 2). Substitution of glycine for each of the five core residues in PU.1, shown on the model, abolishes DNA binding.

framework for the three structural features arranged in a loop-helix-loop pattern that mediate precise DNA binding by the PU.1 domain. In order to delineate the loop-helix-loop motif in other ets domains and to predict whether this motif is the paradigm for ets recognition, we also superimposed the α-carbon skeleton of the fli-1 domain onto the PU.1 backbone bound to the DNA (Fig. 4). Since this is one of an ensemble of structures from the NMR study, detailed comparisons are not possible. However, general comparisons are useful to establish overall structural similarities between the two related molecules. Although the structure of the fli-1-DNA complex was not determined, it should be noted that the published structure of the fli-1 domain (12) reflects a bound conformation since the NMR experiments were conducted on a 98-residue protein frag-

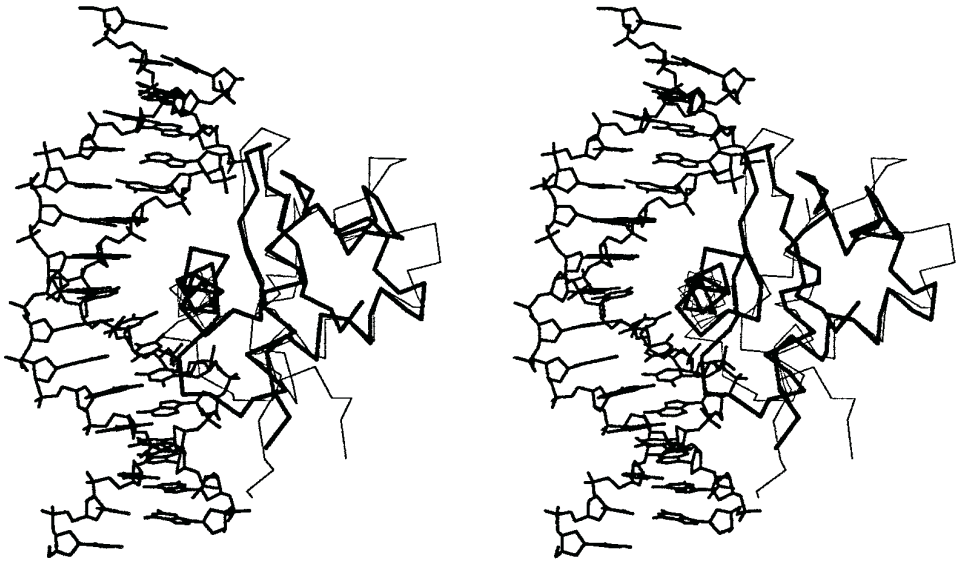

FIG. 4. **Comparison of PU.1 and fli-1 ETS domains.** In this stereo image, the α-carbon backbone of the fli-1 ETS domain (*thin line*; residues 276–373), determined in solution by NMR (12) was superimposed on the PU.1 backbone (*bold line*) using constraints to match similar structural features. The DNA shown in the figure is the oligonucleotide bound to the PU.1 domain.

ment complexed to a 16-base pair oligonucleotide.

As shown in Fig. 4, there is close similarity in the overall scaffold of the ETS domains but several other features of the superposition are worth noting. First, the positions of the four conserved residues that contact DNA are very similar in PU.1 and fli-1. In PU.1, two conserved arginines, 232 and 235, make hydrogen bonds with the bases GGA of the PU core sequence. Arg^{235}(NH-2) forms a hydrogen bond with G-8(O-6) while Arg^{232}(NH-1) makes hydrogen bonds with two bases G-9(O-6) and A-10(N-6) on one strand and a water-mediated contact with T-23(O-4) on the opposite strand. These arginines are strictly conserved in all members of the ets family and the GGA sequence is the consensus DNA sequence recognized by the ets proteins. Therefore, these interactions are expected to be reproduced in all ets protein-DNA complexes. When the fli-1 domain is superimposed on PU.1, the side chains of conserved arginines 232 and 235 in the recognition helix are within hydrogen-bonding distance of the same bases in the GGAA core sequence in the major groove. Substitution of these residues by any other amino acid, even closely related hydrophilic amino acids results in loss of DNA recognition in PU.1, fli-1, and other ets proteins (see Table III). Conserved lysines, residues 219 in the loop (HTH) and 245 in the wing contact the phosphate backbone in PU.1 and are in a position to make the same contacts in fli-1. Mutational substitutions for Lys^{219} in PU.1 (11) and the equivalents of Lys^{219} and Lys^{222} (see Table III) in fli-1 (12) or ets-1 (28) disrupt DNA binding, presumably due to the loss of the phosphate backbone interactions. In fli-1, the equivalents of Lys^{222} and Met^{225} in PU.1 (from the HTH loop) and residues 248/249 (from the wing loop) were identified within 4 Å of DNA by intermolecular NOEs (12). Chemical mapping experiments with the murine ets-1 molecule suggested a similar pattern with a major groove contact zone and interactions with both adjacent minor grooves (30).

DNA Conformation in the PU.1 ETS Domain-DNA Complex—The PU.1 ETS domain contacts DNA over a 10-base pair area. The DNA is bent by 8° in the complex but does not deviate significantly from B-form DNA (see Table II). As can be seen in Fig. 4, the DNA is uniformly curved over the length of the 16-base pair fragment. There is an average helical twist of 33°, with 10.8 base pairs per turn and an average rise per base pair of 3.2 Å. The minor groove is slightly enlarged (~2–3 Å from the mean) in the GGAA region at the midpoint of the oligonucleotide. A "spine" of water molecules, similar to that observed in the crystal structure of a B-DNA dodecamer (31), is located in the minor groove from bases 8 to 12. Binding of the ETS domain induces a DNase I-hypersensitive site 3′ to the C-26 base in the core sequence (30). This site is probably exposed on the face of the DNA opposite to where the protein binds as a result of the expansion of the minor groove (Fig. 3).

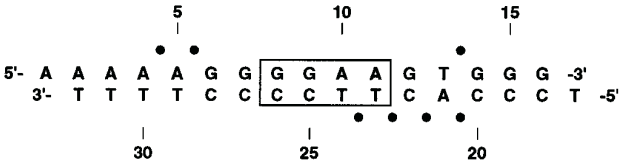

FIG. 5. **Sequence of the oligonucleotide bound to the PU.1 protein in the crystal structure.** The GGAA recognition core sequence as well as the bases on the complementary strand are enclosed in a *box*. The PU.1 domain makes contacts with bases on both strands within this core. The *dots* designate seven phosphates that are neutralized by interactions with basic residues. With the exception of the phosphate at base 14, all of these phosphates lie on one face of the DNA helix.

The DNA bending that is stabilized by the PU.1 domain may serve as an illustration of the hypothesis of DNA bending by phosphate neutralization. It has been demonstrated, by the introduction of neutral methylphosphonate analogues in DNA fragments bearing polyadenylate tracts (32) that bending of the DNA occurs when the phosphate charges are neutralized on one face of the DNA helix, due to repulsion of the remaining anionic phosphates. It was proposed (32) that binding of proteins with cationic surfaces to DNA could also cause the DNA double helix to "spontaneously relax" toward the surface where cationic amino acids neutralized phosphate anions through formation of salt bridges. The PU.1 ETS domain makes neutralizing contacts with phosphate groups on one face of the DNA helix, involving consecutive phosphates on either side of the major groove. The sites of phosphate neutralization are shown on the DNA sequence in Fig. 5. On the GGAA strand, neutralizing contacts with the phosphate backbone 5′ to the core sequence are made by Lys^{208} and Lys^{245} from the wing. On the complementary strand, the phosphate contacts are 5′ to the core sequence as well as with the phosphate backbone within the core: Arg^{173}, Lys^{219}, and Lys^{223} from the HTH loop and Lys^{229} from helix α3. As predicted by the neutralization exper-

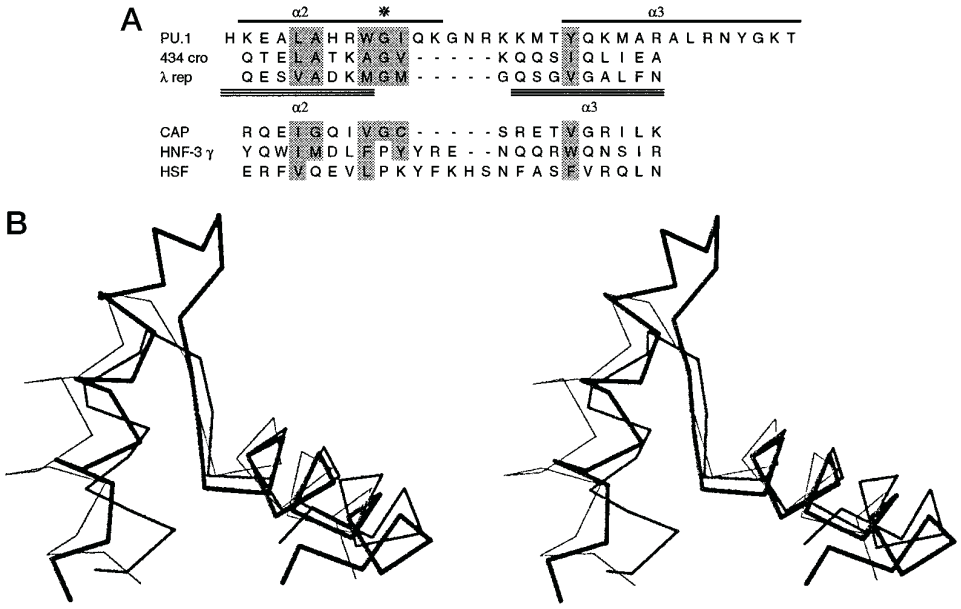

FIG. 6. **Comparison of the HTH motif in PU.1 with other proteins in the HTH superfamily.** *Panel A*, sequence alignment of residues that form the HTH motif in PU.1 with classic HTH proteins 434 cro repressor (42) and λ repressor (41), and with α+β-HTH proteins CAP (15), HNF-3γ (16), and heat shock factor (*HSF*) (40). The structural helices of the PU.1 ETS domain are indicated by *solid bars* above the PU.1 sequence and the helices in the bacterial repressors are indicated by *open bars* below the λ repressor sequence. This figure was adapted from Fig. 1 in Ref. 43. Residues that are *shaded* represent positions in classic HTH that are generally hydrophobic or small (Gly or Ala) in these proteins. The glycine that is conserved in the bacterial HTH proteins is marked with an *asterisk*. Note that helix α2 in PU.1 is one turn longer than the counterpart in the bacterial proteins, yet when the HTH motifs of the repressors are superimposed on the PU.1 HTH, the glycine in the last turn of the PU.1 α2 helix is equivalent to the conserved glycine in the turn of the bacterial proteins (not shown). *Panel B*, the HTH motifs of PU.1 (*thick line*), CAP (medium line; Ref. 15), and heat shock factor (*thin line*; Ref. 40) are superimposed for comparison. The α3 recognition helix is on the *right* in the photograph. Note that the relative orientation of the two helices is closely similar in the three molecules, but the configuration of the residues in the turn between the helices is different. The turn in the PU.1 domain is seven residues in length which is intermediate between the extremes reported for the family of HTH proteins (43, 44).

iments (32), the cationic surface of the PU.1 domain binds to the DNA causing a bend of the duplex oligonucleotide toward the ETS module that is within the range (~10°) of curvature estimated experimentally. The bend is toward the "neutral surface," *i.e.* toward the protein. Two of these phosphate interactions in the minor groove involve conserved residues, Lys^{219} from the HTH loop and Lys^{245} from the wing. Thus the loop-helix-loop pattern may influence both DNA recognition and DNA bending.

This type of charge neutralization is not seen in all protein-induced DNA bends. For example, the TATA-binding protein binds with extensive phosphate backbone interactions to the TATA element (33). Yet in this case the DNA is sharply kinked away from the protein contacts. In CAP (15) salt bridges and other hydrogen bonds to phosphate groups stabilize a severely kinked DNA conformation with DNA bent at 90°.

Interactions with the phosphate backbone are seen in numerous DNA-binding proteins, but these contacts are often hydrogen bonds and not salt bridges. The hypothesis (32) states that neutralization of charge by lysines and arginines results in excess repulsive electrostatic forces that can maintain bending of the DNA double helix (34). The moderate DNA bending seen in the complexes of oligonucleotides with paired homeodomains (35, 36) or HNF-3γ (16) may also result from phosphate neutralization, since these proteins form phosphate-side chain salt bridges with 4 or 3 arginines, respectively. However, the neutralizing contacts are not as extensive as those seen in the PU.1-DNA complex.

The complementarity of the loop-helix-loop motif of fli-1 with the DNA from the PU.1 complex also suggests that, like PU.1, other ETS domains may not significantly deform DNA from B-DNA conformation but to date there is not much biochemical data in the literature on DNA bending by ETS domains. In one study of the ETS domain from the Elk-93 protein, circular permutation analyses indicated that DNA binding by the Elk-93 fragment did not induce significant bending of DNA (37). In contrast, in the human ets-1-DNA complex (14), the DNA was kinked at a 60° angle due to intercalation of a tryptophan side chain. The equivalent of this tryptophan, tyrosine 175 in PU.1, is found in the hydrophobic core and is not in position to intercalate. Substitution of glycine for this tyrosine in PU.1 does not affect DNA binding (11). In fli-1 (12), the equivalent tryptophan is buried in the hydrophobic core and was not listed among residues in close proximity (≤4Å) to DNA. Thus, the molecular basis for kinked DNA cannot be understood in the context of contacts seen in the PU.1-DNA complex (11) or inferred in the fli-1 complex (12). DNA bending by phosphate neutralization is not apparent in the ets-1-DNA complex, since only one lysine and one arginine form phosphate-side chain salt bridges. The arginine is the equivalent of Arg^{235} in PU.1 that forms a hydrogen bond with base G-8 in the GGA core.

Target Specificity—The superimposed models in Fig. 4 suggest that a loop-helix-loop scaffold that brings together conserved amino acids and conserved DNA bases is a general mode of DNA recognition by ets proteins. Yet, ets transcription factors bind to the GGA(A/T) core motif in the context of specific promoters. To begin to identify residues that influence target specificity, it is necessary to look for mutations of non-conserved residues that affect DNA binding. Of the 14 absolutely conserved residues in the domain, seven contact DNA in the PU.1 complex. These contacts would be expected to be maintained for all ets-DNA complexes. In studies of a number of members of the ets family, mutations have been reported that affect DNA binding. These mutations, summarized in Table III, can now be correlated with the atomic model of the PU.1-DNA complex. Some of these residues are conserved residues, but others are unique to a particular molecule.

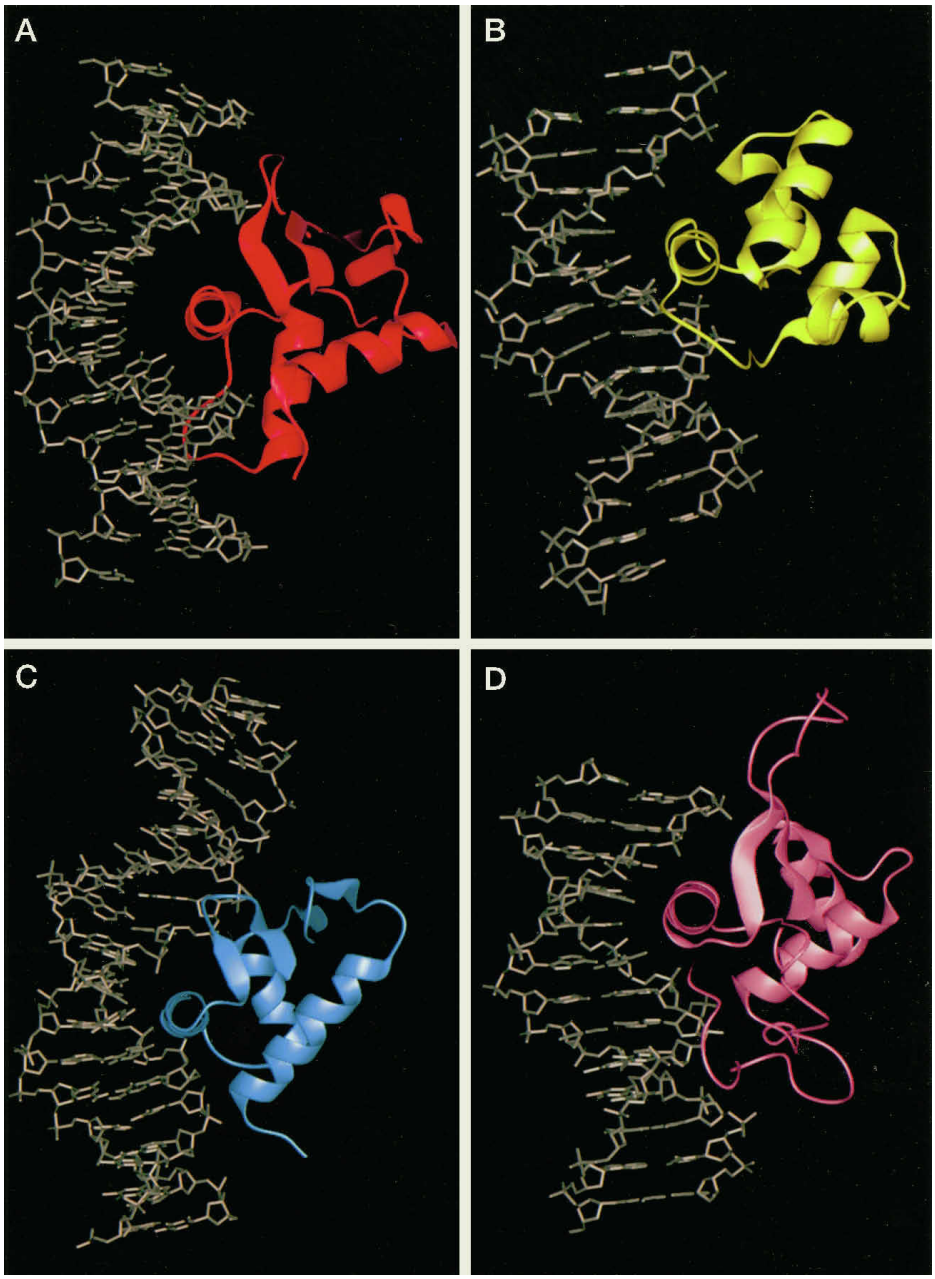

FIG. 7. **Comparison of protein-DNA complexes in HTH proteins.** The loop-helix-loop pattern of DNA recognition in the PU.1 complex (*panel A*) is compared to a classic HTH protein 434 cro repressor (42) (*panel B*), to an $\alpha+\beta$ HTH protein CAP (15) (*panel C*), and to a winged-HTH protein HNF-3γ (16) (*panel D*). In each of these complexes, the recognition helix makes contact in the major groove. The contacts of the PU.1 domain with DNA are more extensive and include interactions from two loops in the minor grooves on either side of the major groove where recognition helix α3 binds.

It should be emphasized that PU.1 contacts both strands at the GGAA core. Interactions are made by conserved residues as well as residues where sequence variability exists in the ets family. Therefore, ets recognition requires specific base contacts with the GGAA sequence and the bases on the complementary strand. For example, it has been shown that a single residue converts DNA recognition of ets proteins from GGAA to GGAT. When a lysine in chicken ets-1 (equivalent to residue 229 in PU.1) is altered to threonine found in this position in Elf-1 and E74, the resultant protein exhibits a restricted selectivity for GGAA like the Elf-1/E74 proteins and the reverse mutation causes the converse change in DNA recognition (38). In the PU.1 complex, Lys^{229} is located in the recognition helix and makes a water-mediated contact to base C-25 on the antisense strand at the GGAA core. There is a water network located in the major groove at the GGAA site. Twelve well defined water molecules are hydrogen-bonded to the bases and also form a hydrogen-bonded network between the two strands. This water network may contribute to the stability of the duplex and consequently influence specific DNA recognition. Since the side chain of lysine is long, it is possible that the contact of a shorter residue such as threonine would not bind to this water network and could contact a different base, *i.e.* T-23. The water network itself could also change. Or, the interchange of lysine↔threonine could permit DNA contact reflecting the stereochemical difference in size of adenine *versus* thymine bases.

HTH Motif—All of the direct contacts with specific bases in the PU.1-DNA complex are made by residues in the α3 recognition helix. Two non-conserved residues, Thr^{226} and Gln^{228}, at the amino-terminal end of this helix, make water-mediated contacts with bases C-25 and C-26, respectively, that are base paired to guanines 8 and 9 in the core GGAA sequence. Both of these residues are unique to PU.1/SpiB in the ets family, so these may represent PU.1-specific contacts.

Tyr^{227}, which is strictly conserved in the ets family, is located in the hydrophobic interior of the protein. While the phenyl ring of this tyrosine is buried, the hydroxyl group is exposed and lies within 3.6 Å of G-6(O1P). This residue was not included in our list of DNA contacts using a conservative cut-off

of 3.2 Å for hydrogen bonds/ionic interactions. Although this interaction may not occur in PU.1, with a simple side chain rotation, a hydrogen bond is possible with the phosphate backbone. This may be an example of a contact made by a conserved residue that influences DNA recognition by selected family members. Substitution of cysteine for this tyrosine abolishes DNA binding in ets-1 (28).

In Fig. 6*A*, the sequence of the HTH motif of PU.1 is compared with the sequence of "classic" bacterial HTH proteins and other winged-HTH proteins. The glycine required in the turn between helices in HTH proteins (39) is also conserved in this position in ETS domains, although the α2 helix is one turn longer than the helix in HTH proteins. In PU.1, the glycine lies in the last turn of this helix. This glycine and other hydrophobic residues in α2 and α3 stabilize the arrangement of these two helices in HTH proteins. Even this pattern of conserved hydrophobic residues is seen in ets proteins. In other winged-HTH proteins, HNF-3γ (16) or heat shock factor (40), the sequence similarities are not as apparent. These two proteins have prolines in the equivalent position of the conserved glycine and the presence of this proline may influence the configuration in the "turn." On the other hand, ets proteins may exhibit a helical arrangement that is structurally closer to that in "classic" HTH proteins. When HTH elements of PU.1 and HTH molecules such as λ (41) or 434 cro (42) repressors are superimposed, the glycine is in a structurally equivalent position (not shown). Moreover, the overall pattern of docking of the recognition helix in the major groove is quite similar when 434 cro repressor (42), CAP (15), and PU.1 are compared bound to DNA (Fig. 7). The major difference is the fact that the recognition helix in PU.1 docks deep in the major groove with contacts to the bases involving residues along the entire length of the helix, while DNA contacts in CAP and other classic HTH proteins are made from residues at the amino-terminal portion of the helix.

None of the related proteins in the HTH superfamily actually contact DNA by residues in the HTH turn (43, 44). This novel DNA contact may be possible in PU.1, as well as other ets proteins, because the connecting segment between helices is more of a loop than a turn. The corresponding HTH motifs of heat shock factor (40) and CAP (15) are compared to PU.1 in Fig. 6*b*. But it is not simply the length of the "turn" or "loop" in the HTH motif that accounts for this DNA contact in PU.1, since other eukaryotic HTH proteins contain even longer connecting segments (43, 44) and yet do not contact DNA by this structural feature, for example HNF-3γ (16). Thus the contacts made by this loop in PU.1 illustrate a new DNA contact that, to date, is unique to the ets proteins as the newest members of the HTH superfamily.

Loops and Minor Groove Contacts—Since the sequences in the HTH loop as well as the loop (wing) between strands β3 and β4 are not strictly conserved among members of the ets family, these residues may be important sites for specific recognition by individual members of the family. In the PU.1-DNA complex, these two loops contact the minor groove through interactions with the phosphate backbone closest to the major groove. It is also interesting to note that the length of both of the contact loops differs among members of the family, with the PU.1 loop containing an "extra" glycine at residue 220 and lacking a glycine after residue 247. Other residues in these loops may also provide specific contacts to bases in other ets proteins. For example, the change of arginine→aspartic acid (equivalent to 244 in PU.1) affects DNA binding in Elk-1 (45).

Since ets proteins bind DNA as monomers, it could be expected that there would be extensive contacts to stabilize the interaction. HNF-3γ also binds DNA as a monomer (16). In the HNF-3γ complex, three regions were involved in DNA recognition: the recognition helix and two wings. The location of the first wing between the last two strands in the β-sheet corresponds topologically to the wing in PU.1, but contacts from the second wing emanate from a loop at the COOH terminus of the domain. The structural equivalent of this second loop is absent in PU.1. In CAP, the major DNA contacts are made from the recognition helix. This protein binds DNA as a dimer. The surface area on CAP that is buried on DNA binding is 1187 Å^2. Similarly, the surface area buried when 434 cro repressor binds DNA is 1306 Å^2. But with the formation of the DNA complex with the PU.1 ETS domain, 1701 Å^2 surface area is buried. The significantly greater surface area of the PU.1 domain covered reflects the extensive protein-DNA contact region extending for more than 30 Å (11).

The PU.1-DNA model suggests that residues from the two loops contribute the critical interactions for recognition of bases other than the conserved GGAA core when the core is embedded in specific promoter sequences. The loops approach segments of the DNA that are adjacent to the conserved core sequence and therefore these interfaces are stereochemically suitable to permit sequence-specific interactions by a given family member while maintaining the consensus interactions at GGA(A/T). Moreover, the contacts from these loops may mediate specific base interactions by stabilizing a bend toward the protein. Future extensive mutational studies of amino acids that contact DNA are needed to identify these residues. Ultimately, crystal structures of other ets proteins complexed to DNA can be compared to distinguish unique DNA contacts.

Acknowledgments—We thank Dr. Roger Fourme and the staff at the LURE synchrotron for consultation and support for this project. In addition, we are grateful to John Knight and Rick Mitchell for synthesis and purification of oligonucleotides, and to Kelly Riddle-Hilde for preparing the manuscript for publication.

REFERENCES

1. Moreau-Gachelin, F. (1994) *Biochim. Biophys. Acta* **1198,** 149–163
2. Wasylyk, B., Hahn, S. L., and Giovane, A. (1993) *Eur. J. Biochem.* **211,** 7–18
3. Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C., and Maki, R. A. (1990) *Cell* **61,** 113–124
4. Watson, D. K., McWilliams-Smith, M. J., Nunn, M. F., Duesberg, P. H., O'Brien, S. J., and Papas, T. S. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 7294–7298
5. Reddy, E. S., and Rao, V. N. (1988) *Oncogene Res.* **3,** 239–246
6. Reddy, E. S., Rao, V. N., and Papas, T. S. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 6131–6135
7. Rao, V. N., Huebner, K., Isobe, M., ar-Rushdi, A., Croce, C. M., and Reddy, E. S. (1989) *Science* **244,** 66–70
8. Wasylyk, C., Kerchaert, J.-P., and Wasylyk, B. (1992) *Genes Dev.* **6,** 965–974
9. Lim, F., Kraut, N., Frampton, J., and Graf, T. (1992) *EMBO J.* **11,** 643–652
10. Petersen, J. M., Skalicky, J. J., Donaldson, L. W., McIntosh, L. P., Alber, T., and Graves, B. J. (1995) *Science* **269,** 1866–1869
11. Kodandapani, R., Pio, F., Ni, C.-Z., Piccialli, G., Klemsz, M., McKercher, S., Maki, R. A., and Ely, K. R. (1996) *Nature* **380,** 456–460
12. Liang, H., Mao, X., Olejniczak, E. T., Nettesheim, D. G., Yu, L., Meadows, R. P., Thompson, C. B., and Fesik, S. W. (1994) *Nature Struct. Biol.* **1,** 871–876
13. Donaldson, L. W., Petersen, J. M., Graves, B. J., and McIntosh, L. P. (1996) *EMBO J.* **15,** 125–134
14. Werner, M. H., Clore, G. M., Fisher, C. L., Fisher, R. J., Trinh, L., Shiloach, J., and Gronenborn, A. M. (1995) *Cell* **83,** 761–771
15. Schultz, S. C., Shields, G. C., and Steitz, T. A. (1991) *Science* **253,** 1001–1007
16. Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993) *Nature* **364,** 412–420
17. Pio, F., Ni, C.-Z., Mitchell, R. S., Knight, J., McKercher, S., Klemsz, M., Lombardo, A., Maki, R. A., and Ely, K. R. (1995) *J. Biol. Chem.* **270,** 24258–24263
18. Leslie, A. (1994) *MOSFLM User Guide,* Mosflm version 5.20, MRC Laboratory of Molecular Biology, Cambridge, England
19. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr.* **50,** 760–764
20. Furey, W., and Swaminathan, S. (1990) *Am. Cryst. Assoc. Meeting* **18,** 73
21. Jones, T. A. (1985) *Methods Enzymol.* **115,** 157–171
22. Brünger, A. T. (1992) *X-PLOR Manual,* Version 3.1, Yale University, New Haven, CT
23. Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T., and Berman, H. M. (1996) *Acta Crystallogr.* **D52,** 57–64
24. Babcock, M. S., and Olson, W. K. (1994) *J. Mol. Biol.* **237,** 98–124
25. Holm, L., and Sander, C. (1994) *Nucleic Acids Res.* **22,** 3600–3609
26. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138
27. Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E. Jr., Brice, M. D.,

Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542
28. Mavrothalassitis, G., Fisher, R. J., Smyth, F., Watson, D. K., and Papas, T. S. (1994) *Oncogene* **9,** 425–435
29. Wang, C.-Y., Petryniak, B., Ho, I.-C., Thompson, C. B., and Leiden, J. M. (1992) *J. Exp. Med.* **175,** 1391–1399
30. Nye, J. A., Petersen, J. M., Gunther, C. V., Jonsen, M. D., and Graves, B. J. (1992) *Genes Dev.* **6,** 975–990
31. Drew, H. R., and Dickerson, R. E. (1981) *J. Mol. Biol.* **151,** 535–556
32. Strauss, J. K., and Maher, L. J., III (1994) *Science* **266,** 1829–1834
33. Kim, J. L., Nikolov, D. B., and Burley, S. K. (1993) *Nature* **365,** 520–527
34. Crothers, D. M. (1994) *Science* **266,** 18–19
35. Xu, W., Rould, M. A., Jun, S., Desplan, C., and Pabo, C. O. (1995) *Cell* **80,** 639–650
36. Wilson, D. S., Guenther, B., Desplan, C., and Kuriyan, J. (1995) *Cell* **82,** 709–719
37. Shore, P., Bisset, L., Lakey, J., Waltho, J. P., Virden, R., and Sharrocks, A. D. (1995) *J. Biol. Chem.* **270,** 5805–5811
38. Bosselut, R., Levin, J., Adjadj, E., and Ghysdael, J. (1993) *Nucleic Acids Res.* **21,** 5184–5191
39. Brennan, R. G., and Matthews, B. W. (1989) *J. Biol. Chem.* **264,** 1903–1906
40. Harrison, C. J., Bohm, A. A., and Nelson, H. C. M. (1994) *Science* **263,** 224–227
41. Beamer, L. J., and Pabo, C. O. (1992) *J. Mol. Biol.* **227,** 177–196
42. Mondragon, A., and Harrison, S. C. (1991) *J. Mol. Biol.* **219,** 321–334
43. Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem.* **61,** 1053–1095
44. Brennan, R. G. (1992) *Curr. Opin. Struc. Biol.* **2,** 100–108
45. Janknecht, R., Zinck, R., Ernst, W. H., and Nordheim, A. (1994) *Oncogene* **9,** 1273–1278
46. Ramachandran, G. N., and Sasiekharan, V. (1968) *Adv. Protein Chem.* **23,** 283–437
47. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Cryst.* **26,** 283–291
48. Soudant, N., Albagli, O., Dhordain, P., Flourens, A., Stéhelin, D., and Leprince, D. (1994) *Nucleic Acids Res.* **22,** 3871–3879