AD_____

GRANT NUMBER DAMD17-94-J-4439

TITLE: Structural Studies of the PU.1 Transcription Factor

PRINCIPAL INVESTIGATOR: Kathryn R. Ely, Ph.D.

CONTRACTING ORGANIZATION: The Burnham Institute
La Jolla, CA 92037

REPORT DATE: October 1996

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1996	Annual (1 Sep 95 - 31 Aug 96)

4. TITLE AND SUBTITLE
Structural Studies of th PU.1 Transcription Factor

5. FUNDING NUMBERS
DAMD17-94-J-4439

6. AUTHOR(S)
Kathryn R. Ely, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The Burnham Institute
La Jolla, CA 92037

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick
Frederick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

Ets transcription factors play a role in development and are implicated in some malignant processes. Recently, members of this large gene family have been identified in normal gene expression in mammary cells and also in breast cancer cell lines. In these studies, the crystal structure of the DNA-binding domain of the PU.1 ets protein complexed to DNA has been determined at 2.3 A resolution. The DNA binding domain is a conserved region that binds the core sequence 5'-GGAA/T-3'. The PU.1 domain binds DNA using a loop-helix-loop motif involving conserved amino acids and bases. In this project we are also using nuclear magnetic resonance (NMR) to determine the unbound structure of the domain in solution. The two structures are being correlated to understand the process of DNA recognition by ets proteins.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	29
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

N/A Where copyrighted material is quoted, permission has been obtained to use such material.

N/A Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

N/A Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

N/A For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Kathryn R. Ely 9/30/96

PI - Signature Date

TABLE OF CONTENTS

Page

Introduction.....5
Body--Progress Report.....6
Task 1.....6
Task 2.....7
Task 3.....8
Task 4.....9
Conclusions.....11
References.....13
Appendix.....14

INTRODUCTION

Transcription factors bind to target DNA sequences to regulate metabolic functions such as growth and differentiation. The PU.1 (spi-1, sfpi-1) transcription factor (1) is a member of the *ets* gene family, a recently discovered family of regulatory proteins. There are now more than 45 members in this family that have been identified in various organisms from *Drosophila* to humans (reviewed in Refs. 2 and 3). These molecules play a role in normal development and have been implicated in malignant processes. Important for these studies is the fact that ets-related proteins have been identified in normal mammary cell-specific gene expression (4) as well as in breast cancer cell lines (5-7).

The ets-related proteins have been proposed to regulate gene expression in mammary tissue and such molecules could influence the production of gene products that are responsible for human breast tumorigenesis (7). It has been shown that elevated expression of the PEA3 gene is directly correlated with the development of metastatic mammary tumors in transgenic mice with the neu oncogene (5). Moreover, in 25-30% of primary human breast cancers, there is an amplification and over-expression of the HER2/neu (cerbB-2) proto-oncogene (6). Overexpression of HER2 is associated with more aggressive tumor growth and reduced patient survival (6). An ets-related response element has been found in the promoter of the HER2 gene and deletion analysis of this promoter revealed that this site is an important *cis*-acting element for HER2 translational activity (7). Thus, an ets protein present in these cells stimulates the expression of HER2 and may be a contributing factor to the development of breast cancers. It has also been found that L-plastin is over-expressed in a number of solid tumors (8), while there are four ets-promoter elements on the L-plastin gene (9). These results, together with those obtained from the study of HER2 expression strongly implicate ets proteins in the development and/or metastatic spread of human breast tumors. Therefore, the structural models that are to be generated in this study will be an important contribution to the future design of therapeutic agents to modulate tissue-specific transcription directed by ets molecules.

The ets proteins share a conserved region of approximately 85 amino acids known as the ETS domain (10) that serves as a DNA-binding domain and recognizes a purine-rich sequence, 5'-C/AGGAA/T-3'. Ets proteins differ

in size and in the relative position of the ETS domain within the intact protein. The remaining sequences in ets proteins are presumed to form other functional domains such as activation domains or inhibitory domains that mask the DNA binding site. The ETS domain binds to DNA as a monomer, unlike many other DNA-binding proteins.

PU.1, the subject of this study, is an ets protein expressed in hematopoietic cells and specifically in immune system cells such as B cells, macrophages, neutrophils or mast cells (1,2). The sequence of PU.1 is identical with the oncogene spi-1 (2). Within the ets family, the PU.1 sequence is the most divergent from ets-1 and yet there is 40% sequence homology in the DNA binding domain of these two proteins (see Figure 1). At the end of the first year of funding, we reported the successful crystallization of the PU.1 ETS domain in complex with DNA. In the second year of funding, we have solved the crystal structure of this complex. This is the first ets protein to be crystallized and, to date, this is the only crystal structure available for an ets protein. We succeeded by varying the length of both protein and DNA fragments in the complex. Last year, we published the methods used for the co-crystallization of the complex (11). Because of the strong sequence homology of the DNA-binding domains, we propose that similar strategies may be useful for crystallization of ETS domains from other members of the *ets* family. In the present annual report, we will describe the crystal structure of the complex and also present significant advances that we have made toward the solution structure by nuclear magnetic resonance studies of the unbound domain.

BODY -- PROGRESS REPORT

During the first 12 months of funding, we accomplished the first two tasks set forth in the statement of work of the original application. The focus of our efforts during months 13 to 24 was primarily on Tasks 3 and 4 and these efforts will be outlined in detail in the following sections.

Task 1. Large scale purification of the PU.1 DNA-binding domain. Months 1-36

Milligram quantities of the DNA-binding domain of PU.1 were expressed in bacteria and purified to homogeneity (11). The protocols developed were scaled up to production level and were standardized and made closely reproducible. These protocols were critical for the production of large crystals. Interestingly, the longer domain fragment that was optimal for crystallization was not ideal for the nuclear magnetic resonance (NMR)

studies in solution. Therefore, we had to produce another protein fragment for the solution studies described under Task 3. Despite implementing different purification protocols for the fragment and verifying that the fragment was intact after time in the NMR tube, the resonance spectra were not complete (see report under Task 3). Crosspeaks were only seen for 94 of the 111 expected backbone amide protons. By this time, from the crystal structure analyses described under Task 4, we already knew that there is inherent flexibility at the amino- and carboxyl-terminal ends of this fragment since there were 11 disordered residues at the amino-terminus and 14 disordered residues at the carboxyl-terminus that were not visible in the electron density map. Therefore we generated a new fragment specifically for the NMR studies that is shorter and expressed from a different vector system (pET3) to optimize yield for labelling. The new fragment is 93 residues in length and includes an N-terminal methionine for bacterial expression. This fragment was purified to homogeneity and tested for stability by gel electrophoresis (see Figure 2). Improved definition of the growth media and culture conditions are now providing consistently high yields (4-6 mg/liter bacterial culture) of soluble protein. More recently, we have produced labelled protein for the high resolution analyses described in Task 3. The efforts to modify the production of protein fragments have been highly critical for the progress that we report for the NMR studies over the the past year.

Task 2. Synthesis of DNA oligonucleotides. Months 1-18

DNA oligonucleotides were screened for binding in complex with the PU.1 domain and those that promoted crystallization of the complex were selected for final screening. Ultimately, a sixteen base-pair oligonucleotide with the sequence 5'-AAAAAGGGGAAGTGGG-3' and the complementary strand were synthesized on the ten micromolar scale, purified by reverse phase HPLC chromatography and annealed. These procedures were optimized to produce fragments that promoted the growth of large crystals of the complex (11). It was evident in the electron density map of the complex that the DNA fragments formed long extended fiber-like elements in the crystal lattice by end-to end-stacking between adjacent oligonucleotides, and that this was a major interaction for the nucleation of crystal growth.

Task 3. Determination of the solution structure of the PU.1 domain by NMR. Months 1-36

With the shorter fragment, prepared specifically for NMR, we have already proceeded directly to experiments with labelled protein for the NMR analysis. Our earlier studies with the longer fragment were slowed because the absence of 17 expected backbone peaks makes the NMR analyses more difficult and ambigous. This new sample should accelerate our progress. The HSQC (Heteronuclear Single Quantum Correlation) spectrum for the 94-residue fragment labelled with isotopic nitrogen is presented in Figure 3. This experiment shows exclusively proton resonances attached to a labelled ^{15}N nucleus. In the spectrum each spot represents a backbone amide proton. This spectrum also shows protons from amino acid side chains that contain nitrogen. For example, the three spots in the lower left of the figure result from tryptophan side chains and the resonances from glutamines and asparagines are indicated in the upper right in the figure. Here we can identify all the expected peaks: 93 observable amide protons from the main chain of the protein, 7 x 2 for the glutamines and asparagine side chains (each side chain has two protons attached to nitrogen), and 3 indole (NH) protons from each of three tryptophans.

These results represent the needed unambigous data essential for the NMR structural studies. The full accounting of all expected amide crosspeaks shown in Figure 3 allows us to proceed directly to more experiments that link all the residues properly in their position and to push towards a tertiary structure of this domain with no gaps in the backbone assignments. Experiments to produce doubly-labelled (^{13}C,^{15}N) protein are underway.

In the process of generating spectra from the short fragment, we have also observed in the initial homonuclear spectra (i.e., protons only) that one of the tryptophan side chains may populate more than one conformation. In the crystal structure, two tryptophans are adjacent in the β-sheet (strand β1). The third tryptophan is located in helix α2 and contacts the DNA backbone in the complex. When glycine is substituted for this tryptophan, DNA binding is lost (12). As an example of how the NMR and crystal structure analyses are complementary, we are testing the effect of temperature and concentration on the "moving" tryptophan. Chemical shift analyses suggest that this residue corresponds to Trp215 that contacts the DNA. We plan to monitor the effects of motion and flexibility of this tryptophan by NMR dynamics. The time frame for this

particular motion and the conformational adjustment upon titration with DNA is an important issue to evaluate the role of this aromatic side chain in DNA recognition. These results have to be confirmed also in the $^{15}N/^{13}C$-labelled sample for proper analysis.

Task 4: Determination of the crystal structure of the PU.1 domain complexed to DNA. Months 1-36

Production of large crystals of the PU.1 ETS domain in complex with a 16 base-pair synthetic oligonucleotide containing the recognition sequence was achieved and reported in last year's report and published (11; reprint included in Appendix). Crystals formed in the space group C2 with *a*=89.1, *b*=101.9, *c*=55.6Å and β=111.2°, with two complexes in the asymmetric unit. Four heavy atom derivatives were prepared by soaking crystals and by co-crystallizing with iodinated oligonucleotides. The locations of the iodinated bases were also used to orient the DNA in the electron density map. Multiple isomorphous replacement phases plus anomalous scattering were used to calculate initial electron density maps at 3 Å resolution. The initial MIRAS map was improved by solvent flattening by the method of Wang (13). The improved electron density map was used to build the model. The density for the DNA helix was a prominent feature of the map. After the DNA was positioned, the polypeptide backbone was fitted with polyalanine and finally side chains were added to the model. There were 11 disordered residues at the amino-terminus and 14 disordered residues at the C-terminus so these amino acids were not included in the model. For all other residues representing the complete ETS domain, the electron density was clear and allowed unambiguous fitting of backbone and side chain atoms. Stereochemistry was optimized to ideal bond and angle parameters in X-PLOR (14). One cycle of simulated annealing was followed by alternate cycles of manual model building and refinement. The structure was reported (12; reprint included in Appendix) at 2.3Å resolution and the high quality of the model was demonstrated with a crystallographic R-factor of 23.7 % and R_{free} of 29.9%.

This is the first crystal structure determined for an ets protein. The PU.1 domain assumes a tight globular structure (33 x 34 x 38 $Å^3$) formed by three α-helices and a four-stranded antiparallel β-sheet (see Figure 4). The domain topology is similar to the structures of other ets family proteins fli-1 (15), murine ets-1 (16) and human ets-1 (17) determined in solution by NMR. The structural studies revealed that ETS domains have a

common folding pattern that is similar to $\alpha+\beta$ helix-turn-helix (HTH) DNA-binding proteins including CAP (18) and resembles 'winged' HTH proteins such as HNF-3γ (19). The domain contacts DNA from three sites: the recognition helix (α3), the loop between β-strands 3 and 4 (a 'wing') and the turn in the HTH motif (α2-turn-α3). This turn is longer than the equivalent in many other HTH proteins and is actually a loop. As shown in Figure 4, the PU.1 domain and probably other members of the ets family, use a DNA-binding motif that can more appropriately be called a loop-helix-loop motif. Therefore our crystal structure reveals a new pattern for HTH recognition and a novel mode of DNA-binding.

Four strictly conserved residues on the surface of the domain are likely to be important for DNA-binding by all members of the ets family. Arg232 and Arg235 emanate from the recognition helix and contact the conserved GGAA sequence in the major groove of DNA. Lys245 from the 'wing' contacts the phosphate backbone of the GGAA strand in the minor groove upstream from the core sequence (see Figure 5) and Lys219 in the loop of the HTH motif forms a salt bridge with the phosphate backbone of the opposite strand downstream of the GGAA core. Substitutions of glycine at each of these four conserved sites abolished DNA binding, confirming the functional importance of these residues. Thus these interactions represent the paradigm for ets recognition which is expected to be reproduced in all ets proteins.

Water molecules also participate in protein-DNA recognition in the PU.1 complex. There are 27 well-defined solvent molecules around the DNA. In the major groove, some of these solvent molecules are hydrogen-bonded to the bases and also form a hydrogen-bonded water network between the two strands that might contribute to the stability of the duplex and influence DNA recognition. The two conserved arginines make water-mediated contacts with the DNA.

The DNA is bent in the complex (8°) when compared to 'canonical' B-DNA structure and is curved uniformly along the entire 16 bp length. The minor groove is slightly enlarged (~2-3Å from the mean) in the GGAA region at the midpoint of the oligonucleotide. Surprisingly, the protein-DNA interactions reported in the NMR structure of a human ets-1-DNA complex (17) differed dramatically from this pattern, involving different contacts and significant (60° kink) DNA deformation. The molecular basis for the kinked DNA cannot be understood in the context of the contacts seen in the PU.1 complex. In the ets-1 structure, this kink is induced by intercalation of the side chain of a tryptophan between bases 6 and 7. The

equivalent of this tryptophan, Tyr175 in PU.1, is located in the hydrophobic core, excluding the possibility for intercalation with the DNA bases. Substitution of glycine for this tyrosine did not affect DNA binding in PU.1. Furthermore, the site of intercalation is located at the opposite extreme of the DNA duplex, upstream of the GGAA core sequence; that is, the ets-1 protein is docked on the DNA 180° from the position of the PU.1 domain. Because of this difference in orientation, in the ets-1 complex, the conserved arginines do not contact the GGAA bases. The striking distinction between the PU.1 and ets-1 structures could reflect extreme evolutionary divergence between members of the ets family although this is highly unlikely. Alternatively, it should be noted that the ets-1 complex was formed under denaturing conditions and it is possible that the Trp intercalation occurred early in the renaturation step with subsequent protein refolding.

Currently, the PU.1 contacts are being tested with mutagenesis to understand these interactions with respect to biological function and other residues are also being substituted to identify residues that mediate recognition of a specific DNA sequence by a given family member.

Conclusions

The work accomplished during the past budget period has been a significant contribution to our understanding of the way that ets proteins recognize DNA. We have successfully produced that first crystal structure of an ets protein and the model will serve as the basis to begin to describe the atomic detail of protein-DNA contacts for this family. During the next budget period, the model will be refined and more solvent atoms will be added. Also, we will extend the resolution by incorporating high resolution data collected at a synchrotron source. We will proceed with the NMR analysis of the unbound domain. The sequential assignment and assignments of the secondary and tertiary structure contacts will be made. In complementary NMR studies, we will emphasize dynamics measurements to assess conformational adjustments that may occur on DNA binding by titration of the PU.1 solution. Besides residues that contact DNA, all other structural elements (helices, loops, etc.) will be monitored by dynamic methods since proteins and protein complexes have intrinsic fluctuations that may influence biological activity. In this manner, the comparison of the structure of the domain complexed with DNA in the crystal and the domain alone in solution by NMR will provide valuable information on the process of DNA recognition by ETS domains. The atomic models will be used to suggest sites of mutation to learn more

about the specific mode of DNA binding by individual family members. The results of mutagenesis on PU.1 as well as other ets proteins will be correlated with the model, highlighting base contacts, phosphate backbone contacts and water-mediated contacts.

1. Klemsz, M.J., McKercher, S.R., Celada, A., Van Beveren, C. and Maki, R.A. (1990) The macrophage and B cell-specific transcription factor PU.1 is related to the ets oncogene. *Cell* **61**:113-124.
2. Moreau-Gachelin, F. (1994) Spi-1/PU.1: an oncogene of the Ets family. *Biochim. Biophys. Acta* **1198**:149-163.
3. Wasylyk, B., Hahn, S.L. and Giovane, A. (1993) The Ets family of transcription factors. *Eur. J. Biochem.* **211**:7-18.
4. Welte, T., Garimorth, K., Philipp, S., Jennewein, P., Huck, C., Cato, A.C.B. and Doppler, W. (1994) Involvement of Ets-related proteins in hormone-independent mammary cell-specific gene expression. *Eur. J. Biochem.* **223**:997-1006.
5. Trimble, M.S., Xin, J.-H., Guy, C.T., Muller, W.J. and Hassell, J.A. (1993) PEA3 is overexpressed in mouse metastatic mammary adenocarcinomas. *Oncogene* **8**:3037-3042.
6. Slamon, D., Godolphin, W., Jones, L., Holt, J., Wong, S., Keith, D., Levin, W., Stuart, S., Udove, J., Ullrich, A. and Press, M. (1989) Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* **244**:707-712.
7. Scott, D.K., Daniel, J.C., Xiong, X., Maki, R.A., Kabat, D. and Benz, C.C. (1994) Binding of an ets-related protein within the DNase-1 hypersensitive site of the HER2/neu promoter in human breast cancer cells. *J. Biol. Chem.* **269**:19848-19858.
8. Lin, C.-S., Park, T., Chen, Z. and Leavitt, J. (1993) Human plastin genes: comparative gene structure, chromosome location, and differential expression in normal and neoplastic cells. *J. Biol. Chem.* **268**:2781-2792.
9. Lin, C.-S., Chen, Z., Park, T., Ghosh, K. and Leavitt, J. (1993) Characterization of the human L-plastin gene promoter in normal and neoplastic cells. *J. Biol. Chem.* **268**:2793-2801.
10. Karim, F., Urness, L., Thummel, C., Klemsz, M., McKercher, S., Celada, A., Van Beveren, C., Maki, R., Gunther, C., Nye, J. et al. (1990) The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence [letter]. *Genes & Dev.* **4**:1451-1453.
11. Pio, F., Ni, C.-Z., Mitchell, R.S., Knight, J., McKercher, S., Klemsz, M., Lombardo, A., Maki, R. and Ely, K.R. (1995) Co-crystallization of an ETS domain (PU.1) in complex with DNA: engineering the length of both protein and oligonucleotide. *J. Biol. Chem.* **270**:24258-24263.
12. Kodandapani, R., Pio, F., Ni, C.-Z., Piccialli, G., Klemsz, M., McKercher, S., Maki, R.A. and Ely, K.R. (1996) A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS domain-DNA complex. *Nature* **380**:456-460.
13. Wang, B.-C. (1985) Resolution of phase ambiguity in macromolecular crystallography. *Meth. Enzymol.* **115**:90-112.
14. Brünger, A.T. (1992) X-PLOR Manual. Yale University, New Haven, CT.
15. Liang, H., Mao, X., Olejniczak, E.T., Nettesheim, D.G., Yu, L., Meadows, R.P., Thompson, C.B. and Fesik, S.W. (1994) Solution structure of the ets domain of Fli-1 when bound to DNA. *Nature* **1**:871-876.
16. Donaldson, L.W., Petersen, J.M., Graves, B.J. and McIntosh, L.P. (1996) Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. *EMBO J.* **15**:125-134.
17. Werner, M.H., Clore, M., Fisher, C.L., Trinh, L., Shiloach, J. and Gronenborn, A.M. (1995) The solution structure of the human ETS1-DNA complex reveals a novel mode of binding and true side chain intercalation. *Cell* **83**:761-771.
18. Schultz, S.C., Shields, G.C. and Steitz, T.A. (1991) Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science* **253**:1001-1007.
19. Clark, K.L., Halay, E.D., Lai, E. and Burley, S.K. (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**:412-420.

APPENDIX

Five figures and two reprints appended

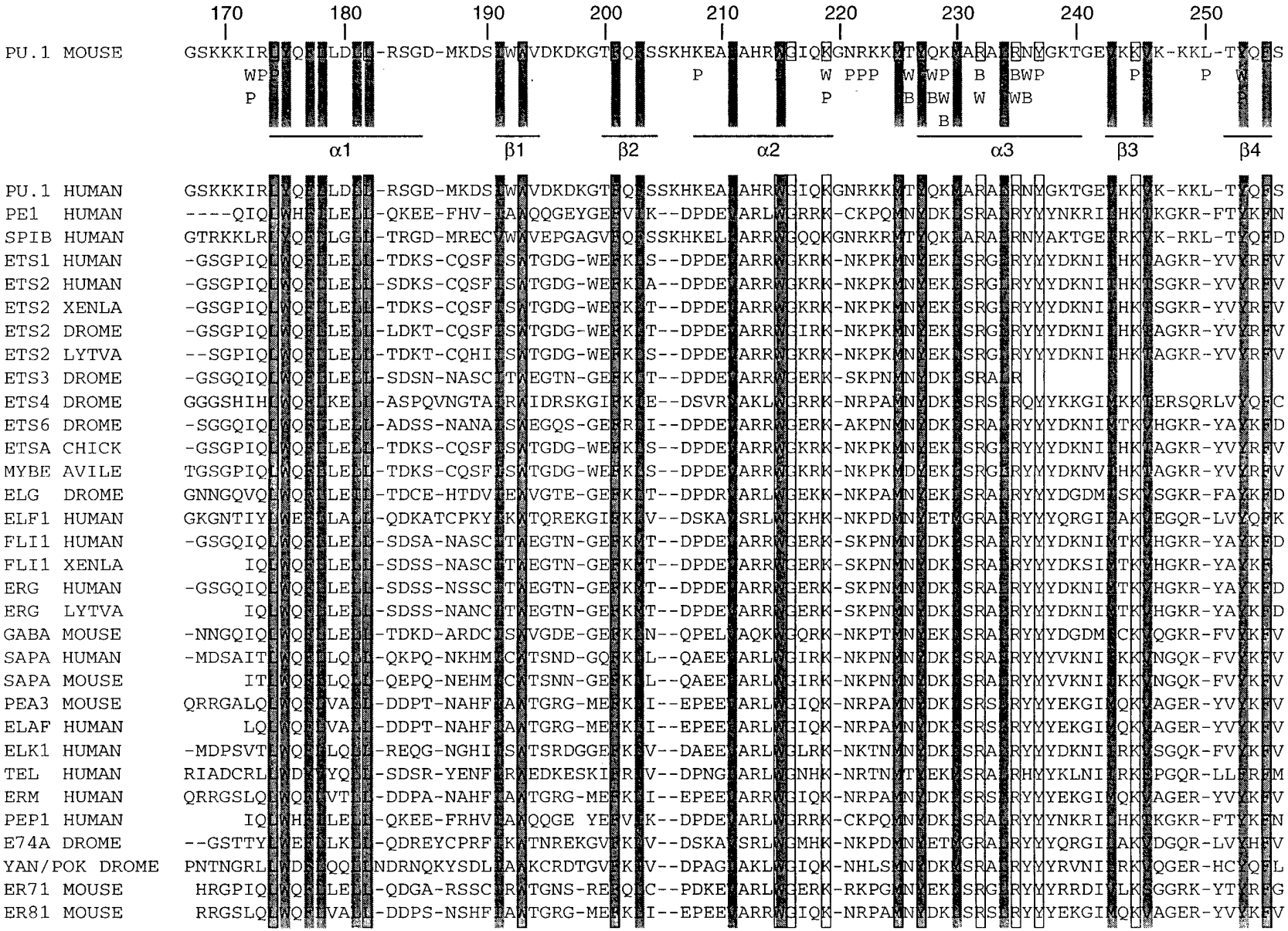
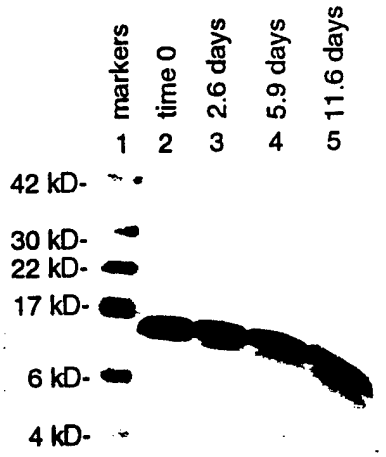
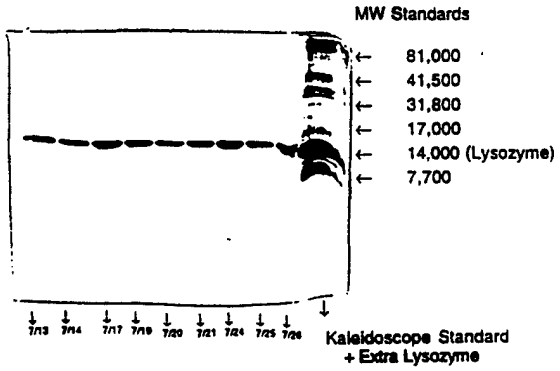

170
180
190
200
210
220
230
240
250
α1
β1
β2
α2
α3
β3
β4
PU.1 MOUSE
PU.1 HUMAN
PE1 HUMAN
SPIB HUMAN
ETS1 HUMAN
ETS2 HUMAN
ETS2 XENLA
ETS2 DROME
ETS2 LYTVA
ETS3 DROME
ETS4 DROME
ETS6 DROME
ETSA CHICK
MYBE AVILE
ELG DROME
ELF1 HUMAN
FLI1 HUMAN
FLI1 XENLA
ERG HUMAN
ERG LYTVA
GABA MOUSE
SAPA HUMAN
SAPA MOUSE
PEA3 MOUSE
ELAF HUMAN
ELK1 HUMAN
TEL HUMAN
ERM HUMAN
PEP1 HUMAN
E74A DROME
YAN/POK DROME
ER71 MOUSE
ER81 MOUSE

Figure 1



Short Fragment



Long Fragment

Figure 2

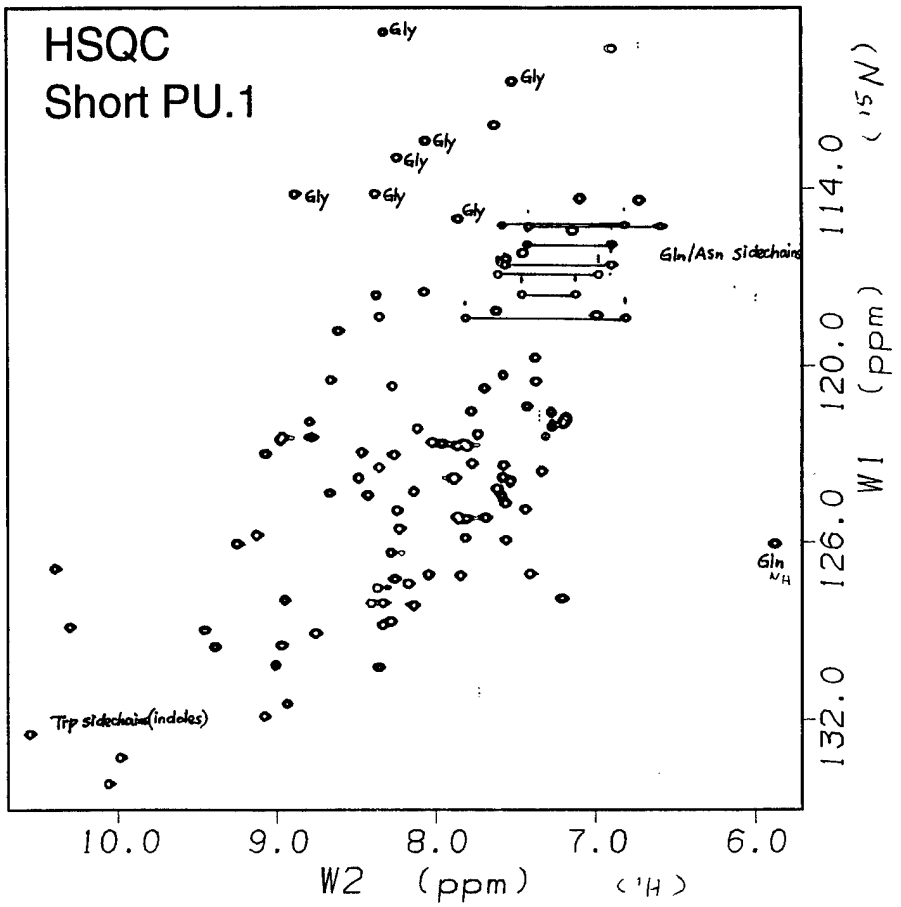


Figure 3

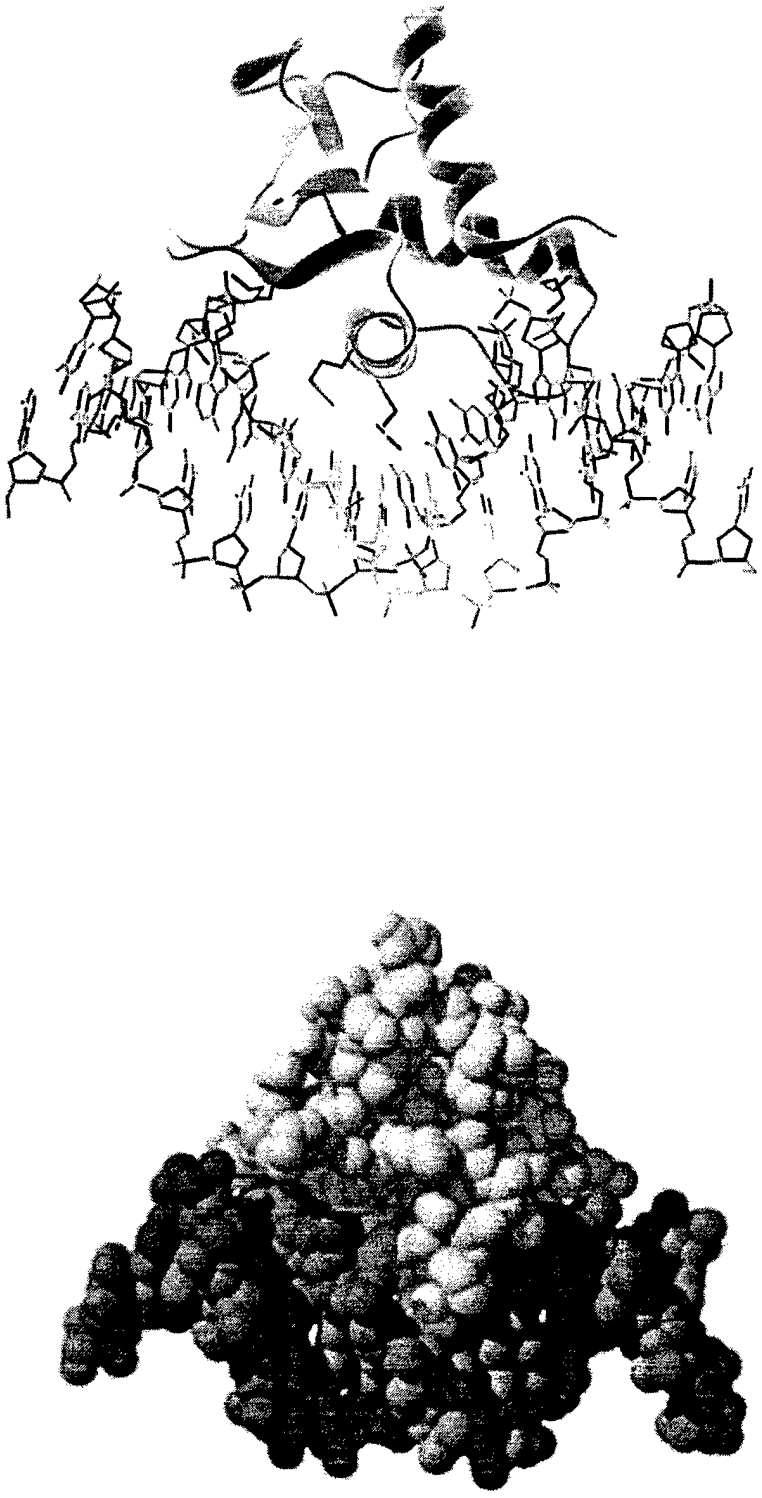


Figure 4

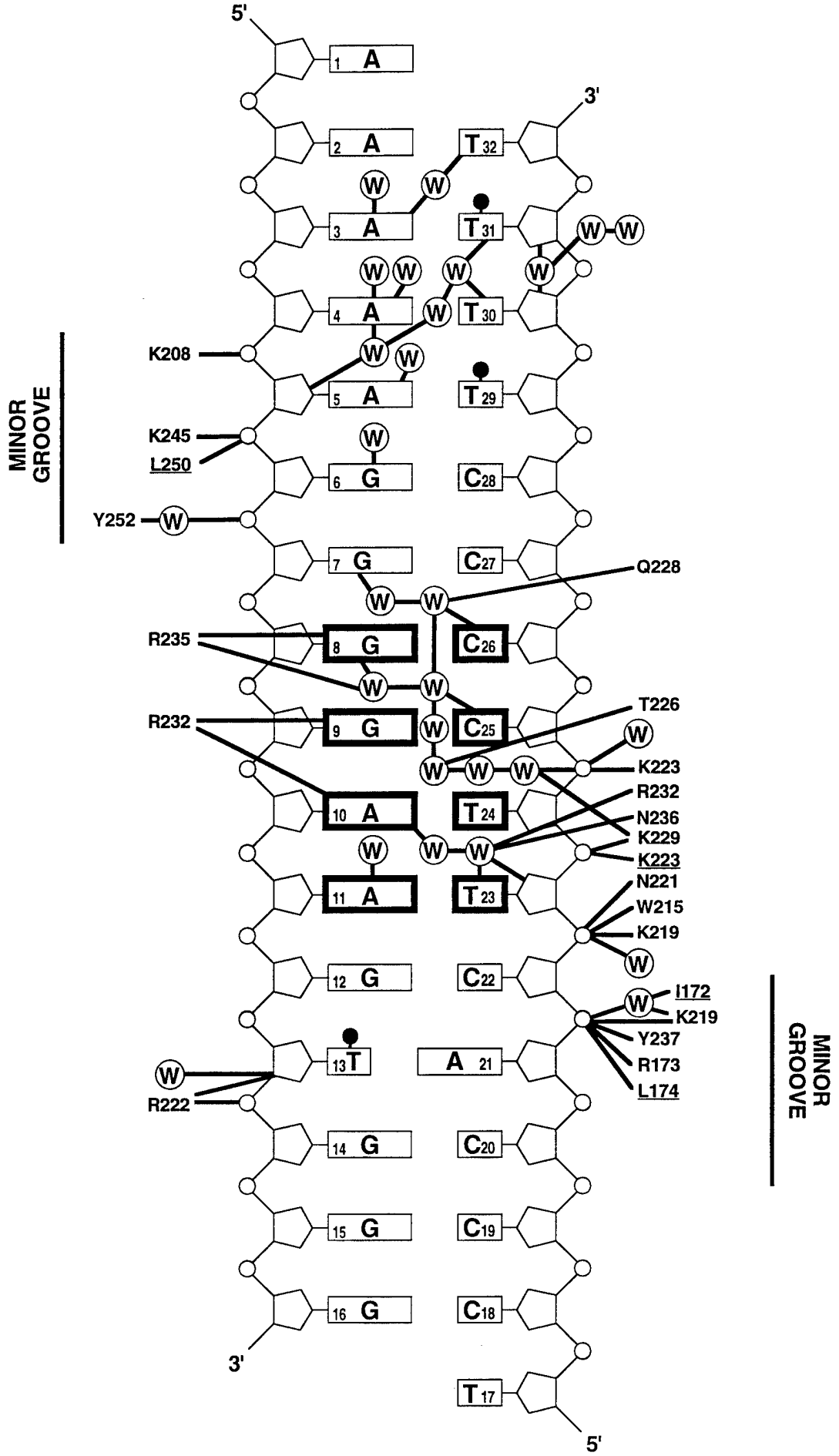

5'
3'
1 A
2 A
3 A
4 A
5 A
6 G
7 G
8 G
9 G
10 A
11 A
12 G
13 T
14 G
15 G
16 G
3'
T32
T31
T30
T29
C28
C27
C26
C25
T24
T23
C22
A 21
C20
C19
C18
T17
5'
MINOR GROOVE
K208
K245
L250
Y252
R235
R232
R222
Q228
T226
K223
R232
N236
K229
K223
N221
W215
K219
I172
K219
Y237
R173
L174
MINOR GROOVE

Figure 5

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 270, No. 41, Issue of October 13, pp. 24258–24263, 1995
Printed in U.S.A.

Co-crystallization of an ETS Domain (PU.1) in Complex with DNA

ENGINEERING THE LENGTH OF BOTH PROTEIN AND OLIGONUCLEOTIDE*

(Received for publication, July 24, 1995, and in revised form, August 9, 1995)

Frédéric Pio, Chao-Zhou Ni, Richard S. Mitchell, John Knight, Scott McKercher, Michael Klemsz‡, Angela Lombardo, Richard A. Maki, and Kathryn R. Ely§

From the Cancer Research Center, La Jolla Cancer Research Foundation, La Jolla, California 92037

The PU.1 transcription factor is a member of the *ets* gene family of regulatory proteins. These molecules play a role in normal development and also have been implicated in malignant processes such as the development of erythroid leukemia. The Ets proteins share a conserved DNA-binding domain (the ETS domain) that recognizes a purine-rich sequence with the core sequence: 5′-C/AGGAA/T-3′. This domain binds to DNA as a monomer, unlike many other DNA-binding proteins. The ETS domain of the PU.1 transcription factor has been crystallized in complex with a 16-base pair oligonucleotide that contains the recognition sequence. The crystals formed in the space group *C*2 with $a = 89.1$, $b = 101.9$, $c = 55.6$ Å, and $\beta = 111.2°$ and diffract to at least 2.3 Å. There are two complexes in the asymmetric unit. Production of large usable crystals was dependent on the length of both protein and DNA components, the use of oligonucleotides with unpaired A and T bases at the termini, and the presence of polyethylene glycol and zinc acetate in the crystallization solutions. This is the first ETS domain to be crystallized, and the strategy used to crystallize this complex may be useful for other members of the *ets* family.

Transcription factors bind to target DNA sequences and regulate important metabolic functions such as cell growth, development, and differentiation. The PU.1 (spi-1, sfpi-1) transcription factor (1) is a member of the *ets* gene family, a recently discovered family of regulatory proteins. There are now more than 35 members in this family that have been identified in various organisms from *Drosophila* to humans (reviewed in Refs. 2 and 3). These molecules play a role in normal development and have been implicated in malignant processes such as erythroid leukemia and Ewing's sarcoma (4). The Ets proteins share a conserved region of approximately 85 amino acids known as the ETS domain (5) that serves as a DNA-binding domain and recognizes a purine-rich sequence with the core sequence, 5′-C/AGGAA/T-3′.

Ets proteins differ in size and in the relative position of the ETS domain. For example, the domain is found near the carboxyl-terminal end of the molecule in PU.1 (Ref. 1; see Fig. 1) and the ets-1 and ets-2 proteins (6, 7), in the middle of the sequence in Erg (8), and within the amino-terminal region in Elk-1 (9). The remaining sequences in Ets proteins are presumed to form other functional domains such as activation domains or inhibitory domains that mask the DNA binding site (10, 11).[1] The ETS domain is sufficient for DNA binding and binds to DNA as a monomer, unlike many other DNA-binding proteins.

Recently, the folding pattern of the DNA-binding domain of Fli-1, an *ets* family protein, was described by NMR analysis (12). The domain consists of 3 α-helices and a four-stranded antiparallel β-sheet. Features of this secondary structure (13) as well as that of the murine ets-1 domain (14) are very similar to the winged helix-turn-helix motif in DNA-binding proteins such as CAP (15) and HNF-3 (16). In order to define precisely the protein-DNA contacts, we co-crystallized the ETS domain of the PU.1 transcription factor in complex with cognate DNA.

The PU.1 transcription factor is expressed in hematopoietic cells and specifically in B cells, macrophages, neutrophils, and mast cells (1, 2). The sequence of PU.1 is identical with the oncogene Spi-1 (17). Spi-1 is activated in the erythroid leukemia induced by spleen focus forming virus. Integration of spleen focus forming virus upstream of the Spi-1/PU.1 gene results in overexpression of the Spi-1/PU.1 protein. This event is associated with the development of erythroid leukemia. The PU.1 molecule has been shown to interact with other nuclear proteins. For example, PU.1 binds to the 3′ enhancer sequence of the Ig-κ gene in complex with a second factor NF-EM5 (PIP) (18, 19). Formation of the ternary complex of PU.1, NF-EM5, and DNA is dependent on PU.1 binding to the core GGAA sequence and phosphorylation of serine 148 in PU.1 (18). The sites of protein-protein interaction and phosphorylation are immediately adjacent and amino-terminal to the DNA-binding domain.

There are several subfamilies of Ets proteins that appear to have arisen by gene duplication of a primordial gene (3). The amino acid sequence of PU.1 is the most divergent from ets-1, yet there is 40% sequence homology in the DNA-binding domains of these proteins. Fourteen residues are strictly conserved in the DNA-binding domain when all ETS domains are compared. Here we report a strategy to clone and express a recombinant fragment encompassing the ETS domain of PU.1 for structural studies. Successful co-crystallization with DNA was dependent on the length of the protein fragment and also on the length of the synthetic oligonucleotide bound to the fragment. It has been shown in studies of other DNA-binding proteins (reviewed in Refs. 20–22) that alteration of the length of DNA oligonucleotides is important to optimize crystallization of the protein-DNA complex. Recently, an extensive analysis of conditions to produce crystals of the U1A-RNA complex was reported (23). In that study, varying the length of RNA

* This work was supported by United States Department of the Army Grant DMD17-94J-4439 (to K. R. E.) and by National Institutes of Health Grant AI20194 (to R. A. M.). 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.

‡ Present address: Dept. of Microbiology and Immunology, Indiana University School of Medicine, Indianapolis, IN 46202-5120.

§ To whom correspondence should be addressed: La Jolla Cancer Research Foundation, 10901 North Torrey Pines Rd., La Jolla, CA 92037. Tel.: 619-455-6480; Fax: 619-455-0181.

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

hairpins as well as utilization of mutant proteins was necessary to produce high quality crystals. The results of the screening of both protein and RNA components were used to propose a general strategy for crystallization of protein-RNA complexes. Since this is the first ETS domain to be crystallized, the details of the selection and production of the protein and DNA components of the complex will be described here. Because of the strong sequence homology of the DNA-binding domains, similar strategies may be useful for successful crystallization of ETS domains from other members of the *ets* family.

MATERIALS AND METHODS

Cloning and Expression of the PU.1 DNA-binding Domain—The DNA-binding domain of PU.1 was cloned in the pET11 expression vector (24) by polymerase chain reaction amplification of the DNA-binding domain from the full-length mouse PU.1 cDNA as described previously (1). DNA sequence analysis was used to verify that the sequence of the amplified product was identical with the original clone. For bacterial expression, pET plasmid constructs were used to transform *Escherichia coli* BL21(DE3)pLysS cells. A preculture of 50 ml of LB medium (25) and ampicillin (100 μg/ml) was inoculated with a single colony from freshly transformed BL21(DE3)pLysS cells bearing the DNA-binding domain insert. After an overnight incubation at 37 °C, this preculture was used to inoculate 7.5 liters of LB-ampicillin media. Cells were grown overnight at 26 °C in an aerated fermentor (Microferm, New Brunswick, NJ). The next morning, 2.5 liters of LB-ampicillin buffered at pH 7.4 with sodium phosphate were added to the culture. Expression of protein was induced at 26 °C with the addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside. After 4 h, cells were harvested by centrifugation and stored as a paste at −70 °C.

Purification of PU.1 DNA-binding Domain—Cell pellets from 1 liter of culture were resuspended in 200 ml of lysis buffer (20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 2 mM EDTA, and 0.1 mM phenylmethylsulfonyl fluoride). Cells were lysed on ice by sonication, cell debris was cleared by centrifugation at 17,000 rpm and 4 °C for 60 min, and the concentration of sodium chloride in the supernatant was adjusted to 1 M. Polyethyleneimine was added to a final concentration of 0.2% and precipitation proceeded with gentle mixing for 30 min on ice. The precipitate was removed by centrifugation at 15,000 rpm and 4 °C for 30 min. The supernatant solution was dialyzed at pH 7.5 against 20 mM Tris-HCl, 60 mM NaCl, and 0.1 mM phenylmethylsulfonyl fluoride and then centrifuged again before application to CM-Sepharose Fast-Flow resin. The ETS domain was isolated by ion exchange chromatography at 4 °C with a linear NaCl gradient (60 mM to 1.2 M). Fractions containing the DNA-binding domain were pooled and concentrated by ultrafiltration. The domain was purified to homogeneity by gel filtration on a Sephacryl S-100 (Pharmacia) molecular sizing matrix at pH 7.4 in phosphate-buffered saline and 0.02% sodium azide. Purified protein was concentrated to 0.5 mM, quick-frozen, and stored in aliquots at −70 °C.

Purification of Selenomethionine-substituted Protein—In order to produce modified protein for structure solution by multiwavelength anomalous dispersion phasing methods (26), recombinant PU.1 DNA-binding domain was produced with selenomethionine substituted for methionine. Bacterial cells (*E. coli* strain B834; Novagen, Inc.) which are auxotrophic for methionine (BL21DE3met-) were used to express the DNA-binding domain. Competent B834 cells were freshly transformed with the pET11 vector containing the domain. For expression of the modified protein, a preculture of 50 ml of LB-ampicillin medium was inoculated with a single colony and incubated at 37 °C. After 16 h, 5 ml of this preculture was used to inoculate 1 liter of M9 medium (25) containing 100 μg/ml ampicillin supplemented with 50 μg/ml selenomethionine (Sigma) and 2 mg/liter each of biotin and thiamine. Cells were grown at room temperature until the absorbance at A_{600} reached 0.15, and expression of recombinant protein was induced by the addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside. After 16 h, cells were harvested by centrifugation and stored at −70 °C. The selenomethionine-substituted protein was purified by procedures described for the native domain. The extent of selenomethionine substitution was evaluated by amino acid analysis and mass spectrometry.

DNA Synthesis and Purification—DNA oligonucleotides of various lengths were synthesized on a 10-μm scale using phosphoramidite chemistry with an Applied Biosystems Model 394 DNA/RNA synthesizer. Derivatized oligonucleotides were synthesized by substituting iodinated uracil phosphoramidites (Glen Research Laboratories) for thymine phosphoramidites. After the last cycle, the oligonucleotides were cleaved from the solid support, and protecting groups on exocyclic amines were removed by treatment with ammonium hydroxide according to manufacturer's protocols before lyophilization. Oligonucleotides were purified by reverse phase HPLC[2] on a Vydac C4 column at 56 °C using an acetonitrile gradient in 100 mM triethylammonium bicarbonate buffer (pH 8.5). Fractions containing the full-length oligonucleotide were pooled and acetonitrile was removed by dialysis against triethylammonium bicarbonate buffer. The oligonucleotides were desalted in 20% ethanol on Bio-Gel P2 resin (Bio-Rad Laboratories, Inc.), lyophilized twice, and stored in aliquots at −70 °C.

Before co-crystallization, DNA extinction coefficients were calculated for each oligonucleotide strand (27), and complementary strands were mixed in equimolar ratios in 5 mM Mes, 200 mM NaCl, pH 7.0, to a final concentration of 0.5 mM. Strands were annealed by heating the mixture to 95 °C and slowly cooling over a few hours to 20 °C.

Space Group Determination and X-ray Data Collection—Crystals were characterized for diffraction using a Rigaku RU-200 rotating anode x-ray source with a graphite monochromator operating at 50 kV and 100 mA, two San Diego Multiwire Systems area detectors, and the UCSD data processing programs (28). Initial characterization and space group determination were performed at room temperature; however, the crystals were sensitive to x-ray exposure. Therefore, all crystals used for this study were cryoprotected in solutions of polyethylene glycol (PEG) and methylpentanediol and immediately frozen in a nylon loop in a cooled nitrogen stream. X-ray data were collected at −145 °C using a cryocooling device and a liquid nitrogen-cooled gas stream (Molecular Structures, Inc.).

RESULTS AND DISCUSSION

Screening of Protein Fragments—Two different recombinant proteins were generated that each include the minimal DNA-binding domain. These fragments are shown in Fig. 1. The two fragments differ in length at both the amino- and carboxyl-terminal ends of the sequence. The amino-terminal sequence and amino acid composition of these fragments indicated that the purified proteins lacked the amino-terminal methionine, probably as a result of proteolytic cleavage by methionyl aminopeptidase (29).

We first generated a protein of 93 amino acids corresponding to residues 168 to 260 since this region encompassed the minimal DNA-binding domain identified by deletion analysis (1). After expression and purification, when this fragment was tested by dynamic light scattering, the protein solution was monodisperse (results not shown) which was a preliminary indication that the recombinant molecule was suitable for crystallization trials (30). However, when the protein was concentrated beyond 5 mg/ml, the fragment formed aggregates and insoluble precipitates. Moreover this fragment was susceptible to proteolytic degradation upon prolonged storage. These observations suggested that the fragment was not folded correctly and that the molecule was not a good candidate for crystallization. After extensive screening, no crystals were obtained with this fragment alone. Only small crystals were observed for this fragment in complex with DNA, and these crystals were difficult to reproduce.

In order to generate a fragment with improved solubility properties, a strategy to alter the length of the molecule was implemented. The design of a construct to produce the longer fragment shown in Fig. 1 was based on secondary structure predictions and an alignment of multiple ETS domain sequences. This analysis indicated that the predicted secondary structure of the sequence at the amino-terminal boundary of the short fragment was not consistent for members of the *ets* family. For PU.1, this region was predicted to form an α-helix, while in the majority of other *ets* family sequences, β-strands were predicted. Therefore, the amino-terminal sequence of the new construct was extended to the boundary of the PEST

[2] The abbreviations used are: HPLC, high performance liquid chromatography; PEG, polyethylene glycol; Mes, 2-(*N*-morpholino)ethanesulfonic acid; bp, base pair(s).

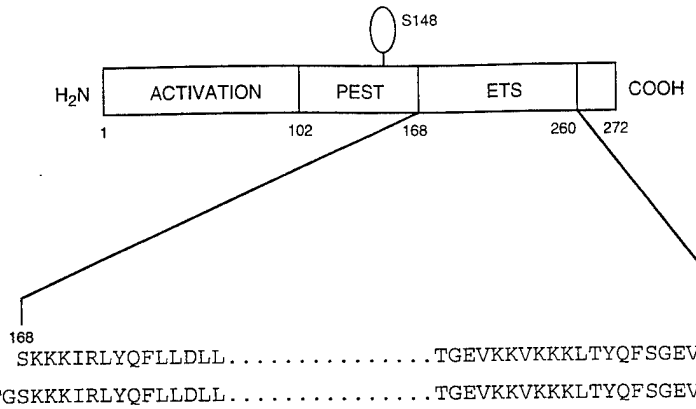


FIG. 1. **Schematic representation of the PU.1 protein.** The sequence of the full-length protein encompasses the activation domain, a PEST region, and the ETS domain which is located at the carboxyl-end of the molecule (reviewed in Ref. 2). The site of phosphorylation (*S148*) that influences protein-protein interactions is labeled (18). Below the molecule, the amino acid sequences for the termini of the two recombinant fragments tested for crystallization are listed. The shorter segment extending from residues 168 to 260 was cloned first; however, this fragment was not a stable protein for structural studies. The longer segment corresponded to residues 160 to 272 which is the actual carboxyl terminus of the full-length PU.1 molecule. This protein was extremely soluble and monodisperse in solution. The amino-terminal serine of this fragment results from the cloning strategy and is not part of the wild-type sequence.

domain excluding a region at the end of the PEST region that is a conserved hydrophilic sequence (see Fig. 1). At the carboxyl terminus, the sequence was extended to the end of the full-length PU.1 molecule. The long fragment encoded by this construct corresponded to residues 160 to 272. After expression and purification, this fragment was remarkably soluble up to concentrations of 60 mg/ml and remained monodisperse in solution even at these high concentrations and after prolonged storage at −70 °C. Despite the optimal physical properties of this fragment, it is surprising that the molecule never crystallized alone even with extensive screening using incomplete factorial (31) and sparse matrix (32) crystallization trials.

Co-crystallization with DNA Oligonucleotides—Some DNA-binding proteins crystallize only when complexed to specific cognate oligonucleotides (reviewed in Refs. 21–22). In many of the complexes crystallized to date, the ends of the DNA fragments interacted in the crystal lattice to form an extended, distorted DNA helix with base-paired interactions between adjacent DNAs in the crystal lattice. In this respect, the oligonucleotides direct the orientation of the complex in the crystal. The PU.1 DNA-binding domain recognizes a purine-rich sequence having a core sequence of 5′-GGAA-3′. The sequences of the oligonucleotides used in this study were identified by screening random sequence oligonucleotides.[1] A number of oligonucleotides were chemically synthesized each of which included the PU box sequence and differed in length. As shown in Fig. 2, oligonucleotides with termini that provide blunt-ended or overhanging bases were tested for co-crystallization.

The quality of the oligonucleotides was critical for successful co-crystallization. In particular, care was taken to achieve >95% homogeneous oligonucleotide by reverse-phase HPLC. The chromatographic separations were run at 56 °C to avoid the formation of secondary structure during purification. Full-length oligonucleotides were eluted from the C4 column with an acetonitrile-triethylammonium bicarbonate gradient. Purification using other gradients or performed on ion exchange resins did not produce oligonucleotides that were adequate for crystallization. After extensive dialysis to remove acetonitrile, each purified oligonucleotide was concentrated by successive lyophilizations from dilute ammonium bicarbonate and was finally desalted in 20% ethanol with a Bio-Gel P2 column. Complete desalting was critical for the formation of large crystals. In fact, DNA heterogeneity or contaminating ions were factors that inhibited crystal growth or produced showers of poorly formed crystals.

Prior to mixing with protein, duplex DNA was annealed by heating to 95 °C and cooling slowly to 20 °C. Molar extinction coefficients were calculated for each strand (22) to ensure that the strands to be annealed were present in equimolar concentrations. Duplex DNA molecules shown in Fig. 2 were mixed with freshly thawed PU.1 protein in molar ratios of 2:1 or 1:1 DNA:protein. In each case, complex formation was verified using a gel shift electrophoretic assay (results not shown). DNA binding was tested with both of the protein fragments. Solubility testing and precipitation analyses were also performed with selected complexes before crystallization trials. The solubility of the protein-DNA complexes was diminished relative to the proteins alone, particularly as compared to the longer PU.1 fragment. In fact, some of the complexes precipitated immediately upon mixing. These precipitates could be redissolved by the addition of NaCl or could be prevented if NaCl was present in the protein solution prior to the addition of DNA. Optimal conditions for mixing PU.1 with DNA were carefully defined yet were dependent on the presence of NaCl at concentrations that varied for each complex.

PU.1-DNA complexes were formed with each of the oligonucleotides shown in Fig. 2 and each of the two PU.1 fragments. Using UV absorbance measurements at 278 nm for protein components and at 260 nm for DNA samples, the final concentration of the complex was estimated at 0.2 mM to 0.4 mM. These complexes were screened for crystallization using the sparse matrix method (32), starting with oligonucleotides >20 bp in length. Trials were set up using vapor diffusion and hanging drops. In these initial screens, crystals grew from conditions that are typical for protein-DNA complexes, *i.e.* neutral pH, polyethylene glycol (PEG), and divalent cations (33).

For complexes with the short protein fragment, only small crystals were obtained in most of the trials. In one case, somewhat larger crystals were observed when the protein was complexed to a 20-bp blunt-ended oligonucleotide, but these crystals could not be improved by complementary screening with shorter oligonucleotides or DNAs with overhanging bases. In contrast, complexes formed with the longer protein fragment were more amenable to screening. The best crystals for this complex initially formed with a 23-bp oligonucleotide with an AT overhang (see Fig. 2). Crystals of this complex were observed in several drops of the screen. The similarity of conditions in each of these trials suggested that sodium acetate was

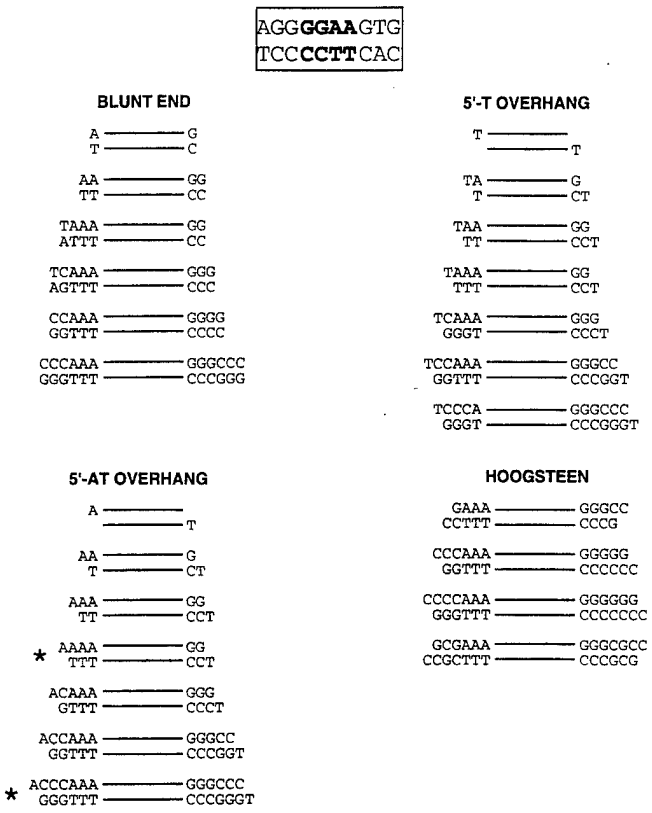


FIG. 2. **Oligonucleotides tested in co-crystallization trials.** Each of the oligonucleotides listed was synthesized for co-crystallization with the PU.1 domain. The sequences differ in length and termini flanking a core sequence shown in the *box* at the *top* of the figure. The core sequence contains the GGAA recognition sequence for PU.1 (*bold*). In each oligonucleotide, the *lines* represent the repetition of this same core sequence. The oligonucleotides were designed to provide both blunt-ended duplex DNA fragments and fragments that have unpaired T or A bases at the termini. The latter were tested because they have the potential for end-to-end stacking in the crystal lattice. The best success with the production of sizable crystals was achieved with two oligonucleotides with a 5'-AT overhang (marked with *asterisks*). The shorter of the two fragments, *i.e.* 16 bp in length, was used to produce diffraction-quality crystals. Other oligonucleotides with unpaired termini were designed to permit Hoogsteen base-pairing between DNA fragments within the crystal lattice. Although the PU.1 DNA binding domain bound these DNA fragments, crystals were never obtained for complexes formed with these oligonucleotides.

essential for crystallization. Tests altering the pH and acetate concentration produced larger crystals of the complex (0.2 × 0.1 × 0.05 mm) after 2 months.

In order to improve these crystals, shorter oligonucleotides were designed. Those with the AT overhang were given priority in the screening. When the long protein fragment was complexed with a 16-bp oligonucleotide with an AT overhang, crystals formed readily as expected; however, under the conditions described above, only crystals with an irregular morphology were obtained. With further screening, well-shaped crystals were produced in drops that contained PEG and zinc acetate. It is interesting that a number of the helix-turn-helix proteins have been crystallized from PEG solutions containing acetate ions. For example, the heat shock factor was crystallized from PEG 4000 and ammonium acetate (34), HNF-3 transcription factor from potassium acetate (without PEG; Ref. 16), NF-κB-50-DNA complex from sodium acetate and PEG 8000 (36), paired homeodomain from ammonium acetate and PEG 1000 (37), and even-skipped homeodomain from potassium acetate and PEG 8000 (38). It appears from this summary that it is a good strategy to test the acetate ion in trials to crystallize helix-turn-helix proteins. Since the presence of zinc acetate produced significant improvement of the PU.1-DNA complex, it is possible that both ions will represent favorable conditions for crystallizing ETS domains. Evaluation of the general utility of these ions awaits the crystallization of other ETS domains.

To our knowledge, this is the first report of a helix-turn-helix protein-DNA complex crystallized in the presence of zinc acetate. In other families of DNA-binding proteins, such as zinc-finger proteins (39) or the diphtheria toxin repressor (40), zinc ions were necessary for crystallization because these molecules have discrete binding sites for the zinc ions in coordination with residues such as histidines or cysteines. In the case of ETS domains, it is possible that the zinc ions also stabilize the protein structure, but identification of the sites for zinc binding awaits the elucidation of the crystal structure.

The PU.1-DNA complex crystals diffracted to 3.5 Å and were improved further by altering the concentration and molecular weight of the PEG used as precipitant. Lower PEG concentrations reduced twinning and excess nucleation. A dramatic improvement in crystal morphology was achieved by substituting PEG 600 for PEG 8000. For the production of large crystals, 5 μl of complex were mixed on a siliconized coverslip with 5 μl of a reservoir solution containing 100 mM sodium cacodylate, pH 6.5, 3–10% PEG 600, and 200 mM zinc acetate. After mixing, the coverslips were inverted and sealed over the reservoir. Parallelopiped crystals formed at 19 °C in 3 to 5 days. In some cases, macroseeding (41) was used to produce large crystals. Crystals were washed free of mother liquor, dissolved, and subjected to nondenaturing gel electrophoresis to confirm the presence of complex.

Diffraction Analyses—These crystals were strongly birefringent and diffracted to at least 2.3-Å resolution. However, the crystals began to dissolve and crack when stored for more than 1–2 weeks and were very sensitive in the x-ray beam. It is interesting that this instability is frequently reported for protein-DNA complex crystals (21). Therefore, crystals were flash-frozen before diffraction experiments in cryoprotectant solutions of 8% PEG 600 and 30% methylpentanediol. A single crystal was quickly transferred from the crystallization drop to the cryoprotectant solution, then picked up in a loop and immediately frozen with a cooled nitrogen stream. After freezing, the crystals were extremely stable in the x-ray beam at −145 °C with no significant decay after 2.5 days of data collection. Flash-freezing did not alter the space group nor significantly change the cell dimensions of the crystals.

The crystals of the PU.1-DNA complex belong to the space group $C2$ with $a = 89.1$, $b = 101.9$, $c = 55.6$ Å, and $\beta = 111.2°$. Assuming a molecular mass for the complex of 22,800 daltons, calculations of the cell dimensions were consistent with V_m (42) of 2.58 Å^3/dalton, solvent content of 48%, and two complexes in the asymmetric unit. These calculations were confirmed by experimental measurements of the crystal density (43). A native data (98% complete) set has been collected at −145 °C to 2.3 Å resolution. The data collection statistics are presented in Table I. The diffraction pattern displayed strong reflections near 3.5 Å that result from scattering of B-DNA which indicated that the DNA oligonucleotides lie approximately along the b axis.

Heavy Atom Searches—Two approaches are being used to obtain heavy atom substitutions for phase calculation. The first approach is to covalently modify the protein and/or DNA components of the complex prior to crystallization and the second is to soak complex crystals in solutions containing heavy metal compounds. In the first strategy, the long PU.1 domain was prepared as a selenomethionine-substituted protein by expression of the recombinant molecule in bacterial culture with

TABLE I
Summary of data collection statistics

Minimum resolution	Average intensity	Average	Number of observations	Number of reflections	R_{sym}[a]
Å	I	$I/\sigma(I)$			
3.93	2,898	48.3	17,522	4,063	0.040
3.12	2,287	36.5	19,299	4,103	0.053
2.73	690	12.1	9,339	4,042	0.079
2.48	405	7.2	7,256	3,969	0.099
2.30	289	4.9	6,679	3,928	0.130
Totals	1,327	22.0	60,095	20,105	0.050

[a] $R_{sym} = \Sigma|I_i - \langle I\rangle|/\Sigma I_i$, where I_i is the intensity of an individual measurement and $\langle I\rangle$ is the mean value of its equivalent reflections.

selenomethionine as the sole source of methionine. There are 3 methionines in the long PU.1 fragment, and substitution of the 3 residues by selenomethionine was confirmed by amino acid analysis (data not shown). The extent of substitution was 70–86% complete in different cultures. The modified protein was co-crystallized in complex with DNA. Large diffraction-quality crystals of this complex were produced that are isomorphous with the native crystals.

In order to modify the DNA for heavy atom substitution, halogenated bases (*i.e.* iodine-substituted uridine for thymine) are suitable for multiple isomorphous replacement methods (*e.g.* Ref. 35). Several iodinated oligonucleotides were synthesized chemically and crystallized in complex with the DNA-binding domain. Iodinated oligonucleotides were tested for binding to the PU.1 molecule by gel shift analyses before co-crystallization. Large isomorphous crystals were obtained with several of these modified oligonucleotides. Besides serving as sites for heavy atom substitution, the iodines may also serve as markers to orient the DNA in the crystal lattice. Since the axis of the DNA is known from the strong reflections in the diffraction pattern, the positions of the iodines at different sites on different oligonucleotides should define the direction of the DNA in the first electron density maps.

Finally, crystals of the native complex are being soaked in heavy atom compounds to produce substitutions for multiple isomorphous replacement phase calculations. Diffraction data for complexes with modified protein and/or DNA are being collected using flash-frozen crystals and ultra-low temperature data collection.

Summary—The production of large diffraction quality crystals of the PU.1 ETS domain in complex with DNA was achieved by a strategy that combined varying the length of both the protein and DNA components of the complex. The DNA fragments used in this study were critical to the successful crystallization for several reasons. Apparently, end-to-end stacking of the oligonucleotides is needed for nucleation of crystal growth since the majority of crystals obtained were from complexes with overhanging bases. Furthermore, the length of the oligonucleotide was important since complexes containing longer oligonucleotides, especially those in the range of 20–23 bp, did not diffract strongly, probably as a result of spacious unoccupied volumes in the crystal lattice. It is interesting that the optimal length for the DNA was 16 bp which corresponds to the length of DNA protected from nuclease cleavage in footprint analyses (1).

While the shorter DNA oligonucleotides were best for crystallization, the longer protein fragment exhibited the ideal physical properties for solubility, DNA binding, and complex crystallization. It is possible that there is an ideal ratio of size of protein to length of DNA for successful crystallization. This ratio relates directly to the shape of the protein component, rather than the oligonucleotide, because the overall shape of the B-DNA is regular and cylindrical. In cases where end-to-end stacking occurs in the crystal, the DNA forms elongated "fiber-like" features arranged side-by-side in the lattice. Since the protein component is usually globular, packing of the bound protein within the lattice formed by neighboring DNA oligonucleotides is important for growth of a three-dimensional crystal. With the parameters reported here and homology-based sequence alignments, it may be possible to design similar protein and DNA fragments to crystallize other ETS domains.

Acknowledgments—We are grateful to Kelly Riddle-Hilde for preparing the manuscript for publication and Muizz Hasham for graphics illustrations. Also, we thank Drs. Craig Hauser and Nuria Assa-Munt for helpful suggestions for bacterial culture and Khanh Nguyen for advice on HPLC chromatography.

REFERENCES

1. Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C., and Maki, R. A. (1990) *Cell* **61,** 113–124
2. Moreau-Gachelin, F. (1994) *Biochim. Biophys. Acta* **1198,** 149–163
3. Wasylyk, B., Hahn, S. L., and Giovane, A. (1993) *Eur. J. Biochem.* **211,** 7–18
4. Hromas, R., and Klemsz, M. (1994) *Int. J. Hematol.* **59,** 257–265
5. Karim, F., Urness, L., Thummel, C., Klemsz, M., McKercher, S., Celada, A., Van Beveren, C., Maki, R., Gunther, C., and Nye, J. (1990) *Genes & Dev.* **4,** 1451–1453
6. Watson, D. K., McWilliams, M. J., Lapis, P., Lautenberger, J. A., Schweinfest, C. W., and Papas, T. S. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82,** 7294–7298
7. Reddy, E. S. P., and Rao, V. N. (1988) *Oncogene Res.* **3,** 239–246
8. Reddy, E. S., Rao, V. N., and Papas, T. S. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 6131–6135
9. Rao, V. N., Huebner, K., Isobe, M., Ar-Rusdi, A., Croce, C. M., and Reddy, E. S. P. (1989) *Science* **244,** 66–70
10. Wasylyk, C., Kerckaert, J.-P., and Wasylyk, B. (1992) *Genes & Dev.* **6,** 965–974
11. Lim, F., Kraut, N., Frampton, J., and Graf, T. (1992) *EMBO J.* **11,** 643–652
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. Liang H., Olejniczak, E. T., Mao, X., Nettesheim, D. G., Yu, L., Thompson, C. B., and Fesik, S. W. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 11655–11659
14. Donaldson, L. W., Petersen, J. M., Graves, B. J., and McIntosh, L. P. (1994) *Biochemistry* **33,** 13509–13516
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. Moreau-Gachelin, F., Ray, D., Mattei, M. G., Tambourin, R., and Tavitian, A. (1989) *Oncogene* **4,** 1449–1456
18. Pongubala, J. M. R., Van Beveren, C., Nagulapalli, S., Klemsz, M. J., McKercher, S. R., Maki, R. A., and Atchison, M. L. (1993) *Science* **259,** 1622–1625
19. Eisenberg, C. F., Singh, H., and Storb, U. (1995) *Genes & Dev.* **9,** 1377–1387
20. Jordan, S. R., Whitcombe, T. V., Berg, J. M., and Pabo, C. O. (1985) *Science* **230,** 1383–1385
21. Harrison, S. C., and Sauer, R. T. (1994) *Curr. Opin. Struct. Biol.* **4,** 1–66
22. Dock-Bregeon, A.-C., and Moras, D. (1992) in *Crystallization of Nucleic Acids and Proteins: A Practical Approach* (Ducruix, A., and Geigé, R., eds) pp. 145–174, IRL Press, New York
23. Oubridge, C., Ito, N., Teo, C.-H., Fearnley, I., and Nagai, K. (1995) *J. Mol. Biol.* **249,** 409–423
24. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1988) *Methods Enzymol.* **185,** 60–89
25. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
26. Hendrickson, W. A. (1991) *Science* **254,** 51–58
27. Puglisi, J. D., and Tinoco, I. J. (1989) *Methods Enzymol.* **180,** 304–325
28. Howard, A. J., Nielsen, C., and Xuong, N. H. (1985) *Methods Enzymol.* **114,** 452–472
29. Hirel, P.-H., Schmitter, J.-M., Dessen, P., Fayat, G., and Blanquest, S. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 8247–8251
30. Ferré-D'Amaré, A. R., and Burley, S. K. (1994) *Bio/Technology* **7,** 1157–1161
31. Carter, C. W., Jr., and Carter, C. W. (1979) *J. Biol. Chem.* **254,** 12219–12223

32. Jancaric, J., and Kim, S.-H. (1991) *J. Appl. Crystallog.* **24,** 409–411
33. Aggarwal, A. K. (1990) in *Methods: A Companion to Methods in Enzymology* (Carter, C. W., Jr., ed) Vol. 1, pp. 83–90, Academic Press, Inc., New York
34. Harrison, C. J., Bohm, A. A., and Nelson, H. C. M. (1994) *Science* **263,** 224–227
35. Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993) *Nature* **363,** 38–45
36. Muller, C. W., Rey, F. A., Sodeoka, M., Verdine, G. L., and Harrison, S. C. (1995) *Nature* **373,** 287–288
37. Xu, W., Rould, M. A., Jun, S., Desplan, C., and Pabo, C. O. (1995) *Cell* **80,** 639–650
38. Hirsch, J. A., and Aggarwal, A. K. (1995) *Proteins* **21,** 268–271
39. Pavletich, N. P., and Pabo, C. O. (1991) *Science* **252,** 809–817
40. Qiu, X., Verlinde, C. L. M. J., Zhang, S., Schmitt, M. P., Holmes, R. K., and Hol, W. G. J. (1995) *Structure* **3,** 87–100
41. Stura, E. A., and Wilson, I. A. (1990) in *Methods: A Companion to Methods in Enzymology* (Carter, C. W., Jr., ed) Vol. I, pp. 38–49, Academic Press, Inc., New York
42. Matthews, B. W. (1968) *J. Mol. Biol.* **33,** 491–497
43. Low, B. W., and Richards, F. M. (1952) *J. Am. Chem. Soc.* **74,** 1660–1666

A new pattern for helix–turn–helix recognition revealed by the PU.1 ETS-domain–DNA complex

Ramadurgam Kodandapani*, Frédéric Pio*, Chao-Zhou Ni*, Gennaro Piccialli†, Michael Klemsz‡, Scott McKercher*, Richard A. Maki*§ & Kathryn R. Ely*

* La Jolla Cancer Research Center at The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA
† Dipartimento di Chimica Organica e Biologica, Universita degli Studi di Napoli Federico II, 80134 Napoli, Italy
§ Neurocrine Biosciences, San Diego, California 92121, USA

THE Ets family of transcription factors, of which there are now about 35 members[1,2], regulate gene expression during growth and development. They share a conserved domain of around 85 amino acids[3] which binds as a monomer to the DNA sequence 5′-C/AGGAA/T-3′. We have determined the crystal structure of an ETS domain complexed with DNA, at 2.3-Å resolution. The domain is similar to α + β (winged) 'helix–turn–helix' proteins and interacts with a ten-base-pair region of duplex DNA which takes up a uniform curve of 8°. The domain contacts the DNA by a novel loop–helix–loop architecture. Four of the amino acids that directly interact with the DNA are highly conserved: two arginines from the recognition helix lying in the major groove, one lysine from the 'wing' that binds upstream of the core GGAA sequence, and another lysine, from the 'turn' of the 'helix–turn–helix' motif, which binds downstream and on the opposite strand.

The PU.1 *[Spi-1, Spfi-1]* transcription factor is an Ets protein expressed in haematopoietic cells[4,5]. PU.1 is a regulatory protein for differentiation of monocytes and macrophages and for B-cell maturation (reviewed in ref. 2). The ETS domain of PU.1 was co-crystallized with a 16 base-pair oligonucleotide containing the recognition sequence[6]. The structure was solved by the multiple isomorphous replacement and anomalous scattering (MIRAS) method (Table 1). The electron density was clearly defined (Fig. 1) for residues 171 to 258, which encompasses the entire conserved ETS domain. The PU.1 domain assumes a tight globular structure ($33 \times 34 \times 38$ Å^3) formed by three α-helices and a four-stranded antiparallel β-sheet (Fig. 1). The domain topology is similar to the structures of other Ets family proteins Fli-1 (ref. 7), murine Ets-1 (ref. 8) and human Ets-1 (ref. 9) determined in solution by NMR. The structural studies revealed a common folding pattern for ETS domains that is similar to α + β helix–turn–helix (HTH) DNA-binding proteins including CAP[10] and resembles 'winged' HTH proteins such as GH5 (ref. 11), HNF-3γ (ref. 12) and HSF (ref. 13). There are three sites of protein–DNA contact: the recognition helix (α3), the loop between β-strands 3 and 4 (a 'wing') and the turn in the HTH motif (α2–turn–α3). The turn between α2 and α3 is longer than the equivalent in many other HTH proteins, and is actually a loop. The DNA-binding motif in PU.1, and probably other members of the Ets family, can be described more appropriately as a loop–helix–loop motif. Therefore the large Ets family defines a new variant subclass of the helix–turn–helix DNA-binding proteins with a novel mode of DNA recognition.

The protein–DNA contacts in the PU.1 complex are detailed in Fig. 2. Four strictly conserved residues on the surface of the domain are likely to be important for DNA binding by all members of the Ets family. Arg 232 and Arg 235 emanate from helix α3 and contact bases in the GGAA sequence in the major groove. These contacts represent the core structure for DNA

‡ Present address: Department of Microbiology and Immunology, Indiana University School of Medicine, Indianapolis, Indiana 46202-5120, USA.

recognition by members of the Ets family because they involve both strictly conserved amino acids and bases in the consensus sequence recognized by these transcription factors (see Fig. 3*b*). The equivalent arginines 81 and 84 in Ets-1 (ref. 9) do not contact the GGAA bases, but intermolecular nuclear Overhauser effects between these arginines and DNA were observed in the Fli-1 NMR studies[7]. Lys 245 extends from β3 just adjacent to the loop ('wing'), and Lys 219 is located in the 'loop' of the HTH motif. Lys 245 contacts the phosphate backbone of the GGAA strand in the minor groove upstream from the core sequence (Fig. 3*c*) and Lys 219 forms a salt bridge with the phosphate backbone of the opposite strand downstream of the GGAA core (Fig. 3*d*). Substitutions of glycine at each of these four conserved sites abolished DNA binding, confirming the functional importance of these contacts (see Fig. 2).

Mutations of conserved residues that contact the phosphate backbone also affect DNA binding. Substitution of glycine at Leu 174 or Trp 215 abolished DNA binding in PU.1. Similarly, substitution of any amino acid in Ets-1 (ref. 14) at the equivalent of PU.1 residues Lys 219 and Arg 222 that bind the phosphate backbone disrupted DNA binding. These minor-groove contacts might represent a conserved pattern for protein 'docking' in the Ets family. In Fli-1 (ref. 7), the equivalents of Leu 174, Lys 219 and Lys 222 showed large chemical shifts on DNA binding in the NMR studies (the counterpart of Trp 215 was buried).

Water molecules also participate in protein–DNA recognition in the PU.1 complex (Fig. 2). There are 27 well-ordered solvent molecules around the DNA. Solvent molecules in the major groove are hydrogen-bonded to the bases and also form a hydrogen-bonded network between the two strands that might contribute to the stability of the duplex and consequently influence specific DNA recognition. Conserved Arg 232 and Arg 235 each form direct and water-mediated contacts with the bases. Three other residues also contact DNA bases through water molecules: Thr 226, Gln 228 and Asn 236. These residues are not conserved in the Ets family and might represent interactions that are unique to the PU.1 protein. Thr 226 and Gln 228, at the amino-terminal end of helix α3, make water-mediated contacts with bases C25 and C26 respectively that are base-paired to guanines 8 and 9 in the core sequence.

TABLE 1 Structure determination and refinement

	Native	Hg	*I* (29)	*I* (13)	*I* (31)
Phasing statistics					
Resolution (Å)	2.3	3.0	2.9	3.0	2.8
Observed reflections	60,095	25,081	20,709	20,512	23,308
Unique reflections	20,105	14,902	13,258	12,910	15,397
Completeness (%)	97	79	65	69	68
R_{sym} (%)*	5.0	3.6	4.0	4.3	3.6
R_{iso} (%) to 3.0 Å†		13.0	14.4	15.9	13.0
Number of sites		2	2	2	2
For isomorphous data ($I/\sigma \geq 3$)					
Phasing power‡		1.33	1.76	1.04	0.98
To resolution (Å)		3.0	3.0	3.0	3.0
R_{Cullis}§		0.62	0.57	0.68	0.67
For anomalous data ($I/\sigma \geq 3$)					
Phasing power		1.0	1.41	1.13	1.43
To resolution (Å)		3.0	3.0	3.0	3.0
Mean figure of merit (10–3.0 Å) is 0.65.					
Refinement statistics					
Resolution range		8–2.3 Å			
Average *B* (Å^2)		20.1			
Crystallographic *R*-factor (%)		23.7			
R_{free} (%)[16]		29.9			
Number of reflections used		16,898 $F > 3\sigma(F)$			
Number of protein atoms		1,486			
Number of DNA atoms		1,300			
Number of solvent atoms		88			

The crystallization of the PU.1 ETS domain (residues 160–272) with a 16-bp synthetic DNA oligonucleotide containing the recognition sequence was described previously[6]. Crystals formed in the space group C_2 with $a = 89.1$, $b = 101.9$, $c = 55.6$ Å and $\beta = 111.2°$, with two complexes in the asymmetric unit. **Phase determination.** Four heavy-atom derivatives were prepared by soaking crystals of the native complex and by co-crystallizing iodinated oligonucleotides with the PU.1 domain. The locations of the iodinated bases are indicated in Fig. 2. Multiple isomorphous replacement phases, including anomalous data, were calculated. The package PHASES[17] was used to refine heavy-atom positions, *B*-factor/occupancies and to calculate phases to 3.0-Å resolution with an overall figure of merit of 0.65. The initial MIRAS map (3.0 Å) was improved by solvent flattening by the method of Wang[18] and with non-crystallographic density averaging. **Model building and refinement.** The improved MIRAS electron-density map was used to build the model with the interactive graphics programs TOM based on FRODO[19] and O[20]. The density for the DNA helix was a prominent feature of the map. To fit the DNA, an 'ideal' B–DNA duplex was generated with the program QUANTA (Molecular Simulations, Inc.) and fitted to the density as a rigid body. After the DNA was positioned, a polyalanine chain was constructed with the BONES option of the Alberta/Caltech program TOM. Subsequently side chains for all residues with clear electron density were added to the model. There were 11 disordered residues at the N terminus of the domain and 14 disordered residues at the C terminus so these amino acids were not included in the model. For all other residues representing the complete ETS domain, the electron density was clear (see Fig. 1) and allowed unambiguous fitting of both backbone and side-chain atoms. Manual adjustments of individual DNA bases were made to fit the electron density. In the program X-PLOR[21], the stereochemistry of the protein was optimized to bond and angle parameters developed by Engh and Huber[22] and for DNA by using parameters of Parkinson *et al.*[23]. Weak restraints were placed on all ribose conformations. One cycle of simulated annealing at 3,000 K (ref. 24) was followed by cycles of manual model building, positional refinement and *B*-factor refinement. More data were added as the refinement progressed in increments: 3, 2.8, 2.6 and 2.3 Å. A total of 88 solvent oxygens ($\langle B \rangle = 22$ Å^2) have been added to the model at this stage of the refinement. Main-chain torsion angles for all non-glycine residues fall within energetically favourable Ramachandran boundaries[25]. The r.m.s. difference for 84 α-carbon atoms in the two complexes in the asymmetric unit is 0.35 Å.

* R_{sym} is $\sum |I - \langle I \rangle| / \sum \langle I \rangle$.

† R_{iso} is $\sum ||F_{PH}| - |F_P|| / \sum |F_P|$, where $|F_P|$ and $|F_{PH}|$ are structure factors for the protein and derivative, respectively.

‡ Phasing power is the r.m.s. value of $|F_H|/E$, where E is residual lack of closure.

§ $R_{Cullis} = \sum ||F_{PH}| \pm |F_P| - |F_{H(calc)}|| / \sum |F_{PH} - F_P|$ for centric reflections, where $F_{H(calc)}$ is the calculated heavy-atom structure factor.

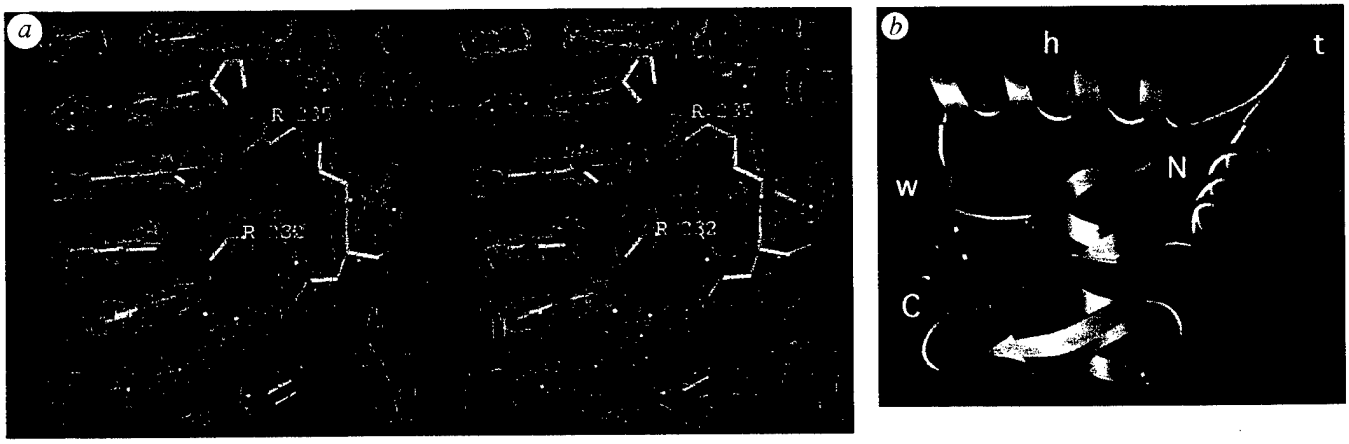


FIG. 1 Overall structure of the PU.1–DNA complex. *a*, Stereoview of the refined 2.3-Å (1.5σ) $2|F_o| - |F_c|$ electron-density map at the protein–DNA interface. Interactions of protein (gold), DNA (red) and water (white) in the major groove at the GGAA core recognition sequence are shown. The two strictly conserved residues, Arg 232 and Arg 235 from recognition helix α3, make direct contact with bases in the major groove. There is a tight network of water molecules at this site in the major groove (Fig. 2). *b*, Ribbon drawing of the PU.1 ETS domain. The module (residues 171–258) is composed of three α-helices and a four-stranded antiparallel β-sheet. In the interior of the domain, a hydrophobic core is formed with 19 side chains including seven strictly and eight highly conserved residues. The major structural features that contact the DNA are indicated: the recognition helix α3 (h), the turn in the HTH motif (t) and the loop between β-strands 3 and 4 (w) corresponding to the 'wing' in these proteins. At the N-terminal end of the fragment, helix α1 begins at residue 172. The C-terminal segment, which is disordered in the PU.1–DNA complex, assumes an α-helical conformation in the unbound Ets-1 NMR structure[8]. This segment might unfold in PU.1 with DNA binding. *c*, Space-filling model of the PU.1 ETS domain–DNA complex. Protein–DNA interactions include both major and minor groove contacts over a distance of 30 Å. The PU.1 transcription factor (gold) binds to DNA as a monomer, so it is not surprising that extensive DNA contact sites exist in addition to the recognition sequence to stabilize binding. HNF-3γ (ref. 12) and GH5 (ref. 11) also bind to target DNA as monomers. In the HNF-3γ–DNA 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 C terminus of the domain. The structural equivalent to the second 'wing' is absent in PU.1.

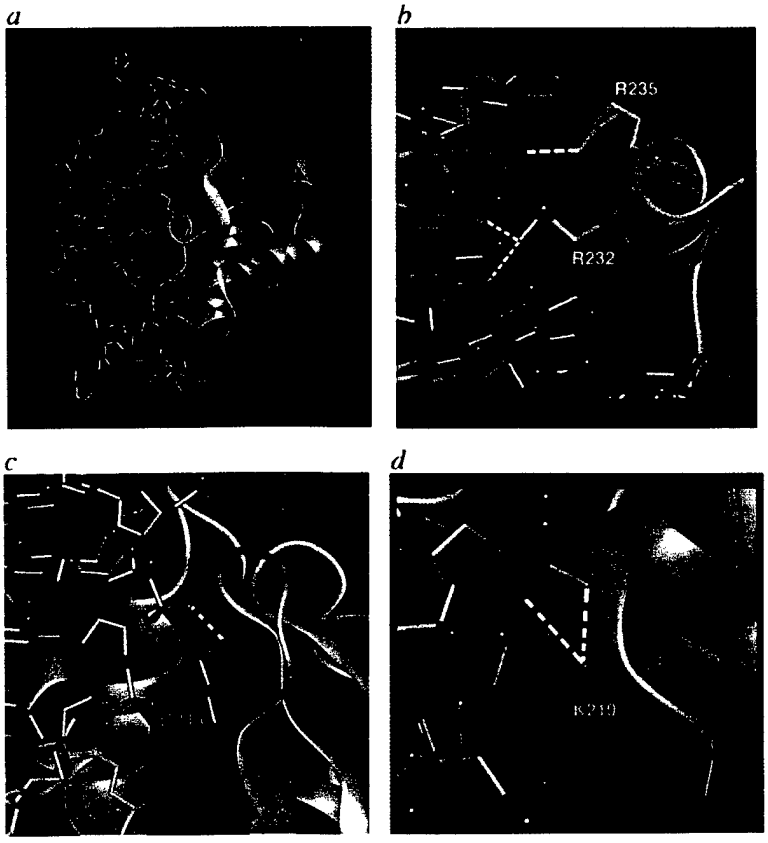


FIG. 3 PU.1–DNA complex. *a*, The 16-bp oligonucleotide bound in complex to the PU.1 ETS domain is shown in grey, with the GGAA sequence coloured red. The ETS domain is represented by an orange ribbon model with the side chains for four conserved residues that contact DNA shown. When glycine was introduced at each of these sites, DNA binding was lost (Fig. 2). *b*, Detailed close-up view, showing that Arg 232 and Arg 235 from the recognition helix make hydrogen bonds with the bases GGA of the PU core sequence. Arg 235(NH2) forms a hydrogen bond with G8(O6), whereas Arg 232(NH1) makes hydrogen bonds with two bases G9(O6) and A10(N6). 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 the interactions shown here represent the paradigm for Ets recognition, which is expected to be reproduced in all Ets protein–DNA complexes. *c*, Interaction of the 'wing': Lys 245(NZ) contacts the phosphate backbone at G6(O2P). *d*, Interaction of Lys 219(NZ) from the loop in the HTH motif, which contacts the phosphate backbone at C22(O3P) and T23(O2P). This figure was generated with the graphics program QUANTA (Molecular Simulations, Inc.).

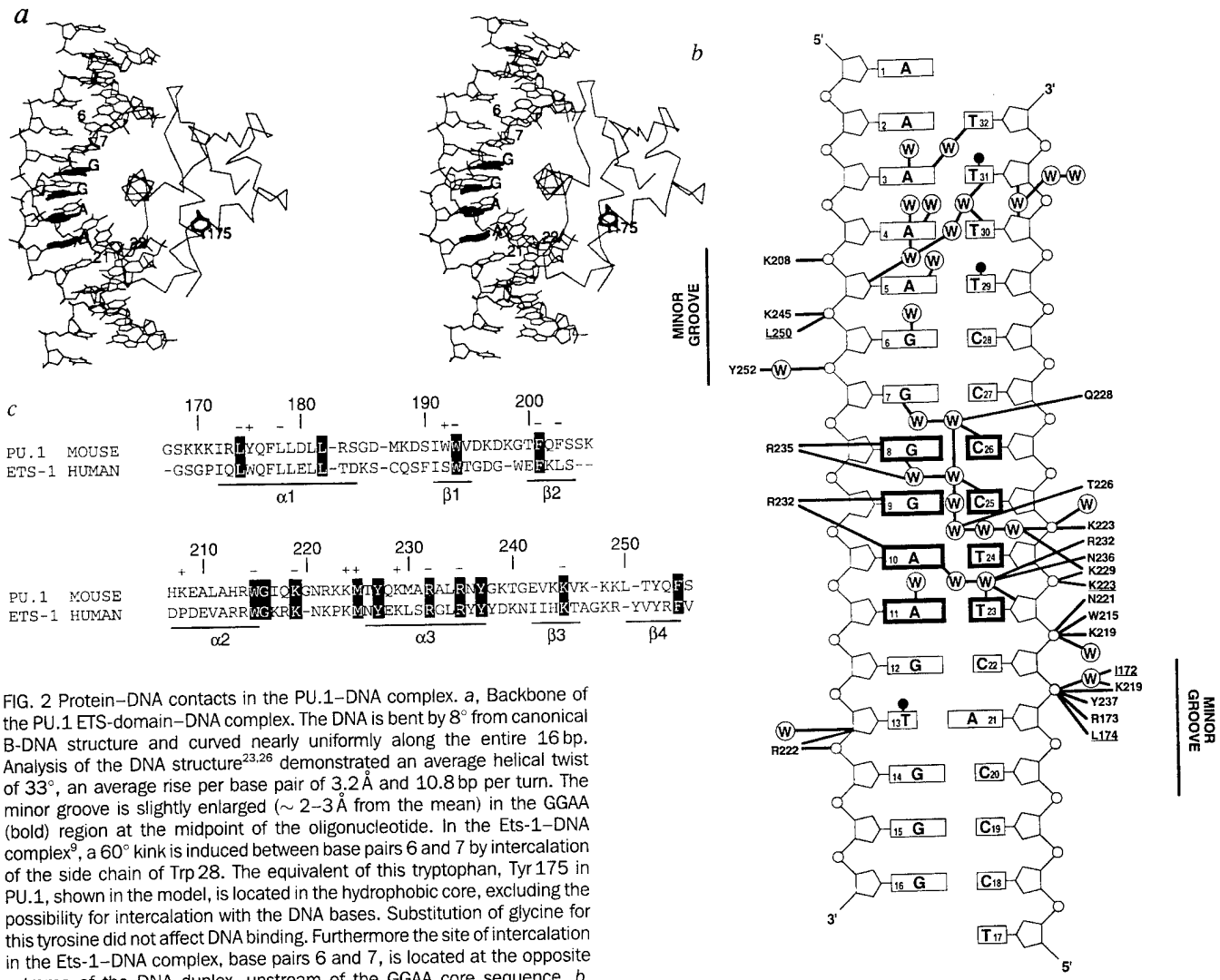


FIG. 2 Protein–DNA contacts in the PU.1–DNA complex. *a*, Backbone of the PU.1 ETS-domain–DNA complex. The DNA is bent by 8° from canonical B-DNA structure and curved nearly uniformly along the entire 16 bp. Analysis of the DNA structure[23,26] demonstrated an average helical twist of 33°, an average rise per base pair of 3.2 Å and 10.8 bp per turn. The minor groove is slightly enlarged (~ 2–3 Å from the mean) in the GGAA (bold) region at the midpoint of the oligonucleotide. In the Ets-1–DNA complex[9], a 60° kink is induced between base pairs 6 and 7 by intercalation of the side chain of Trp 28. The equivalent of this tryptophan, Tyr 175 in PU.1, shown in the model, is located in the hydrophobic core, excluding the possibility for intercalation with the DNA bases. Substitution of glycine for this tyrosine did not affect DNA binding. Furthermore the site of intercalation in the Ets-1–DNA complex, base pairs 6 and 7, is located at the opposite extreme of the DNA duplex, upstream of the GGAA core sequence. *b*, Sequence of the oligonucleotide bound to the PU.1 protein (GGAA PU box in bold lines). Residues that contact the DNA through main-chain atoms are underlined. Well-defined solvent molecules located within 3.2 Å of protein or DNA atoms are identified by an encircled W. Contacts from residues of the 'wing' are made with the nucleotides upstream of the GGAA sequence, and residues from the loop in the HTH motif interact with the opposite strand, downstream of the GGAA site. The direction of the DNA was confirmed by the location of the three iodinated bases (13,29,31; black dots) used for phase calculation. Seven of the residues that contact DNA are strictly conserved and four others are highly conserved. *c*, Sequence alignment of the PU.1 and Ets-1 ETS domains, representing extremes of evolutionary divergence in the family. Residues strictly conserved in all Ets proteins are shown in black boxes; dashes indicate gaps within the family. Numbering and secondary structural features of the PU.1 domain are indicated. The results of mutational analysis when glycine was substituted for a residue are also shown. The effects of the interchanges are labelled + or − above the sequence, indicating that DNA binding was retained or abolished. Mutations were generated essentially as described[27].

The turn in the HTH motif is actually a loop, and because the sequences in this loop as well as the loop ('wing') between strands β3 and β4 are not strictly conserved among members of the Ets family, these residues might be important sites for specific recognition by individual members of the family. In fact, the lengths of both of the contact loops differ among members of the family, with the PU.1 loop containing an 'extra' glycine at residue 220 and lacking a glycine after residue 247. Such conformational differences are expected between family members, but the contrast between the PU.1 and Ets-1 complexes was unexpected. The striking distinction in the mode of DNA contact by the PU.1 and Ets-1 domain could reflect extreme evolutionary divergence between members of the Ets family. Alternatively, it should be noted that the Ets-1–DNA complex was formed under denaturing conditions[9,15] and it is possible that the Trp intercalation occurred early during the renaturation step with subsequent protein refolding.

Future extensive mutational studies of amino acids that contact DNA in Ets proteins are needed to identify residues that mediate recognition of a specific DNA sequence by a given family member. Ultimately, crystal structures of other Ets proteins complexed to DNA must be compared to distinguish unique DNA contacts. □

Received 12 January; accepted 19 February 1996.

1. Wasylyk, B., Hahn, S. L. & Giovane, A. *Eur. J. Biochem.* **211,** 7–18 (1993).
2. Moreau-Gachelin, F. *Biochim. biophys. Acta* **1198,** 149–163 (1994).
3. Karim, F. *et al. Genes Dev.* **4,** 1451–1453 (1990).
4. Klemsz, M. J., McKercher, S. R., Celada, A., Van Beveren, C. & Maki, R. A. *Cell* **61,** 113–124 (1990).
5. Moreau-Gachelin, F., Mattei, M. G., Tambourin, R. & Tavitian, A. *Oncogene* **4,** 1449–1456 (1989).
6. Pio, F. *et al. J. biol. Chem.* **270,** 24258–24263 (1995).
7. Liang, H. *et al. Nature struct. Biol.* **1,** 871–875 (1994).
8. Donaldson, L. W., Petersen, J. M., Graves, B. J. & McIntosh, L. P. *EMBO J.* **15,** 125–134 (1996).
9. Werner, M. H. *et al. Cell* **83,** 761–771 (1995).
10. Schultz, S. C., Shields, G. C. & Steitz, T. A. *Science* **253,** 1001–1007 (1991).

11. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L. & Sweet, R. M. *Nature* **362,** 219–223 (1993).
12. Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. *Nature* **364,** 412–420 (1993).
13. Harrison, C. J., Bohm, A. A. & Nelson, H. C. M. *Science* **263,** 224–227 (1994).
14. Mavrothalassitis, G., Fisher, R. J., Smyth, F., Watson, D. K. & Papas, T. S. *Oncogene* **9,** 425–435 (1994).
15. Werner, M. H., Clore, G. M., Gronenborn, A. M., Kondoh, A. & Fisher, R. J. *FEBS Lett.* **345,** 125–130 (1994).
16. Brünger, A. T. *Nature* **355,** 472–475 (1992).
17. Furey, W. & Swaminathan, S. *Am. crystallogr. Ass. Meeting* **18,** 73 (1990).
18. Wang, B. C. *Methods Enzymol.* **115,** 90–112 (1985).
19. Jones, T. A. *Methods Enzymol.* **115,** 157–171 (1985).
20. Jones, T. A., Zhou, J., Cowan, S. W. & Kjeldgaard, M. *Acta crystallogr.* **A47,** 110–119 (1992).
21. Brünger, A. T. *X-PLOR Manual* Version 3.1 (Yale Univ. Press, New Haven, CT, 1992).
22. Engh, R. A. & Huber, R. *Acta crystallogr.* **A57,** 392–400 (1991).
23. Parkinson, G., Vojtechovsky, J., Clowney, L., Brünger, A. T. & Berman, H. M. *Acta crystallogr.* **D52,** 57–64 (1996).
24. Brünger, A. T., Krokowski, A. & Erickson, J. W. *Acta crystallogr.* **A46,** 585–593 (1990).
25. Ramachandran, G. N. & Sasiekharan, V. *Adv. Protein Chem.* **23,** 283–438 (1968).
26. Babcock, M. S. & Olson, W. K. *J. molec. Biol.* **237,** 98–124 (1994).
27. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. *Gene* **77,** 51–59 (1989).

ACKNOWLEDGEMENTS. This work was supported by grants from the U.S. Army and NIH. R.K. and F.P. made equal contributions to this study. We thank J. Knight and R. Mitchell for synthesis and purification of DNA oligonucleotides, M. Hasham for excellent graphics illustrations, and K. Riddle-Hilde for preparing the manuscript for publication.

CORRESPONDENCE AND MATERIALS. Requests to be addressed to K.R.E.