AD_____

GRANT NO: DAMD17-94-J-4439

TITLE: Structural Studies of the PU.1 Transcription Factor

PRINCIPAL INVESTIGATOR(S): Kathryn R. Ely, Ph.D.

CONTRACTING ORGANIZATION: La Jolla Cancer Research Foundation
La Jolla, California 92037

REPORT DATE: September 1995

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, 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
	September 1995	Annual 1 Sep 94 - 31 Aug 95

4. TITLE AND SUBTITLE
Structural Studies of the 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)
La Jolla Cancer Research Foundation
La Jolla, California 92037

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, 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 words)

Transcription factors bind to target DNA sequences and regulate important metabolic functions such as cell growth, development and differentiation. There is great interest in characterizing these regulatory proteins, especially when their role is linked to oncogenesis. The PU.1 transcription factor is a member of the *ets* gene family and is identical to the *spi.1* oncogene. The PU.1 factor and other ets molecules have been implicated in oncogenic processes such as the development of erythroid leukemia. Recent studies suggest that ets proteins contribute to tumorigenesis in breast cancers. The ets proteins share a conserved region of approximately 85 amino acids (ETS domain) that serves as a DNA-binding domain and recognizes a purine-rich sequence with the core sequence: 5'-GGAA/T-3'. In this project, x-ray crystallography and nuclear magnetic resonance (NMR) are used to determine the structure of the ETS domain of PU.1 alone and complexed to DNA. Atomic models will be used to examine sites in the protein that mediate specific DNA recognition. The structure of the domain will be compared to other DNA-binding motifs to understand the molecular basis for the function of *ets*-related transcription factors and to evaluate the role of these regulatory elements in the development of leukemia, breast cancer, and other cancers.

14. SUBJECT TERMS
transcription factor, x-ray crystallography, multidimensional NMR, DNA-binding domain, protein-DNA complex
breast cancer

15. NUMBER OF PAGES
50

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

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet *optical scanning requirements*.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	PR	- Project
G	- Grant	TA	- Task
PE	- Program Element	WU	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- **DOD** - See DoDD 5230.24, "Distribution Statements on Technical Documents."
- **DOE** - See authorities.
- **NASA** - See Handbook NHB 2200.2.
- **NTIS** - Leave blank.

Block 12b. Distribution Code.

- **DOD** - Leave blank.
- **DOE** - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- **NASA** - Leave blank.
- **NTIS** - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

FOREWORD

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

NA Where copyrighted material is quoted, permission has been obtained to use such material.

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

NA 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.

NA 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).

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

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

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

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

Accession For	
NTIS GRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist | Avail and/or Special	
A-1	

Kathryn R. Ely 9/21/95
PI - Signature Date

Table of Contents

INTRODUCTION

The members of the ETS-domain family of DNA-binding proteins are related to each other by a high degree of sequence similarity within an 85 amino acid segment, which is the DNA-binding domain (1). There is considerable interest in ets proteins because a number of them have been linked to oncogenic processes. The *PU.1 (Spi-1,Spfi-1)* gene (2,3), which is the subject of this study, has been implicated in the development of murine erythroid tumors induced by Spleen Focus Forming Virus (SFFV). Integration of SFFV upstream of the *Spi-1/PU.1* gene results in over-expression of the Spi-1/PU.1 protein. This event is associated with the development of erythroid leukemia.

In very interesting recent results, several laboratories have demonstrated that *ets* transcription factors may contribute to tumorigenesis in breast cancer (4-7). It has been shown that elevated expression of the *ets*-related *PEA3* gene is directly correlated with the development of metastatic mammary tumors in transgenic mice with the *neu* oncogene (4). Moreover, in 25-30% of primary human breast cancers, there is an amplification and overexpression of the *HER2/neu (c-erb-2)* protooncogene (5). Overexpression of *HER2* is associated with more aggressive tumor growth and reduced patient survival (5). 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 transcriptional activity (6). Thus, an ETS-domain protein, present in these cells, stimulates the expression of HER2 and may be a contributing factor to the development of breast cancers. The gene for L-plastin, which encodes an actin-binding protein and is normally expressed only in hematopoietic cells, is activated in a number of solid tumors. A survey of human tumor cell lines revealed a high level of L-plastin in mammary carcinomas (8). Analysis of the promoter of the L-plastin gene revealed four ets-1 responsive elements (9), and it has been suggested by the authors of this study that an ETS-domain protein may be responsible for the abnormal expression of L-plastin in these tumors. These results, together with those obtained from the study of HER2 expression, strongly implicate ETS-domain proteins in the regulated expression of genes that are overexpressed in human breast cancer.

There are now more than 35 members of the *ets* family of transcription factors that have been identified in various organisms from *Drosophila* to humans. 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 (2) and the ets-1 and ets-2 proteins (10,11), in the middle of the sequence in erg (12), and within the amino-terminal region in elk-1 (13). 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 (14,15; Klemsz and Maki, personal communication). The ETS domain is sufficient for DNA binding and binds to DNA as a monomer, unlike many other DNA-binding proteins. The core sequence recognized by ets proteins is: 5'-C/AGGAA/T-3'.

Recently, the folding pattern of the DNA-binding domain of fli-1, an ets family protein, was described by NMR analysis (16). The domain consists of three α-helices and a four-stranded antiparallel β-sheet. Features of this secondary structure (17) as well as

that of the murine ets-1 domain (18) are very similar to the winged helix-turn-helix motif in DNA-binding proteins such as CAP (19) and HNF-3γ (20). No crystal structure has yet been determined for an ets-related protein. Moreover, the mode of DNA contact for the ets proteins remains, for the most part, uncharacterized. In the fli-1 structural studies, intermolecular NOEs between ^{13}C-labelled protein and unlabelled DNA indicated that seven residues were within 4 Å of the DNA and the results suggested that helix α3 was the recognition helix. In order to precisely define the protein-DNA contacts, we proposed to co-crystallize the PU.1 ETS domain with cognate DNA and to determine the structure of the unbound domain in solution by NMR. These structures will provide insight into the active configuration of this transcription factor. In addition, if there are conformational changes in the protein (or DNA) on binding, these differences will be defined in the detailed comparison of the domain alone and in the complex with DNA.

BODY--PROGRESS REPORT

The experimental plan for these structural studies was outlined in a statement of work in the original application and our progress for Months 1-12 will be reported relative to the tasks and timetable projected in the statement of work. As will be described in detail in the following sections, we are proceeding with the experiments on schedule or, in some aspects of the work, well ahead of schedule. The goals in Tasks 1 and 2 have essentially been achieved and the protocols are clearly established to produce milligram quantities of highly purified protein and DNA oligonucleotides for the structural studies. The success of the entire project depends on these procedures, so our progress in these two tasks represents a significant accomplishment that bears directly upon the future progress of the remaining period of support. Also, the fact that the protein and DNA components can be prepared reproducibly with strict quality control is critical for continuity with samples used for data collection in experiments that are performed months apart during the study.

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

To produce large quantities of the protein for structural studies, the DNA-binding domain of PU.1 was cloned in the pET11 expression vector (21) by PCR amplification from the full-length mouse PU.1 cDNA (2). The recombinant domain was expressed in *E. coli* BL21(DE3)pLysS. Bacterial cultures were scaled up to 7-10 liter cultures, and the expression of the recombinant domain was induced by the addition of IPTG. Cells were harvested by centrifugation and then lysed by sonication. Lysates containing the recombinant domain were applied to a CM-Sepharose ion-exchange column and the ETS domain was eluted with a linear NaCl gradient. The domain was purified to homogeneity by gel filtration.

Two different recombinant proteins were generated that each encoded the minimal DNA-binding domain (see Figure 1). The two fragments differed in length at both the N- and C-terminal ends of the sequence. We first generated a protein of 93 amino acids corresponding to residues 168-260 since this region encompassed the minimal

DNA-binding domain identified by deletion analysis (2). However, this fragment tended to form aggregates and insoluble precipitates when concentrated beyond 5 mg/ml for the structural studies. When tested for crystallization in extensive screens, no crystals were obtained with this fragment alone. When tested for crystallization in complex with DNA oligonucleotides, only small crystals were observed and these crystals were difficult to reproduce.

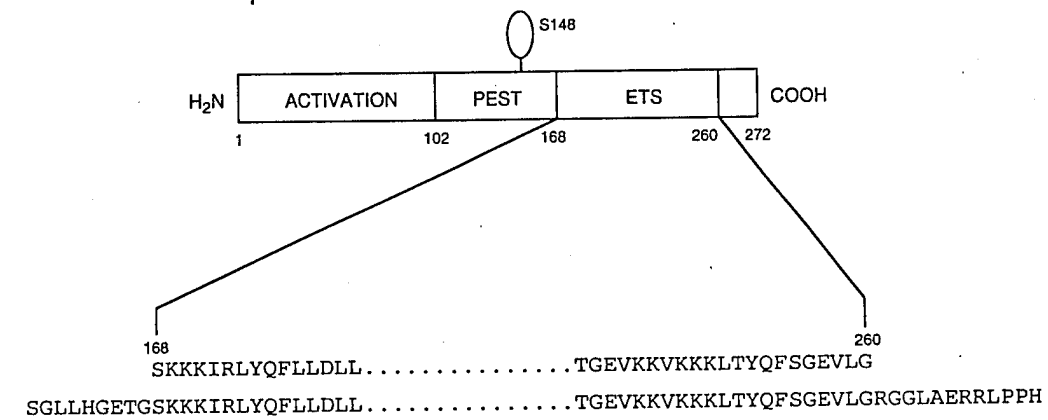


Figure 1. Schematic representation of the PU.1 protein. The sequence of the full-length protein encompasses the activation domain, a PEST region and the ETS domain which is located at the carboxyl-end of the molecule. The amino acid sequences of the termini of the two recombinant fragments generated in this study are listed. The longer fragment was extremely soluble and is being used for crystallography and NMR.

In order to produce a fragment with improved solubility properties, a strategy to alter the length of the molecule was implemented. The design of the longer construct in Figure 1 was based on secondary structure predictions of homologous ets proteins. The N-terminal sequence was extended to the boundary of the PEST domain excluding a segment at the end of the PEST region that is a conserved hydrophilic sequence. At the C-terminal end, the sequence was extended to the end of the full-length PU.1 molecule. The longer fragment (residues 160-272) was expressed in bacteria, purified and was remarkably soluble. The fragment was monodisperse in solution when tested by dynamic light scattering, an early indication that the fragment would be ideal for structural studies. This fragment was then produced in milligram quantities for both crystallization and NMR studies.

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

DNA oligonucleotides are being synthesized on the 10 μM scale for the structural studies using standard phosphoramidite chemistry. The quality of the oligonucleotides is critical for the structural studies, so we have developed protocols specifically to maximize the purification of the synthetic DNA fragments. After the last cycle, the oligonucleotides are cleaved from the solid support and the protecting groups removed before lyophilization. Care is taken to achieve >95% homogeneous oligonucleotide by reverse-phase HPLC, and the separations are run at 56°C to prevent the formation of secondary structure during the purification. Full-length oligonucleotides are eluted with an acetonitrile-ammonium bicarbonate gradient.

Each oligonucleotide is dialyzed and concentrated by successive lyophilizations from ammonium bicarbonate and finally desalted in ethanol on a Biogel P2 column.

The sequences of the oligonucleotides used in this study were identified by screening random oligonucleotides (Klemsz and Maki, personal communication). A number of oligonucleotides were synthesized that each included the PU box core sequence and differed in length (see Figure 2), including those with termini to provide blunt-ended or overhanging bases. Some DNA-binding proteins only crystallize when complexed to specific cognate oligonucleotides. In many such complexes, the ends of the DNA fragments interact 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 packing, or at least the orientation, of the complex in the crystal lattice. Our goal was to drive the crystallization through selection of the optimal length of oligonucleotide for the complex. Therefore, each of the oligonucleotides shown in Figure 2 was synthesized on the large scale, purified and tested in DNA binding gel shift assays for complex formation with the PU.1 ETS domain. The domain bound to each of these DNA fragments and consequently, each of the oligonucleotides were tested in co-crystallization with the domain.

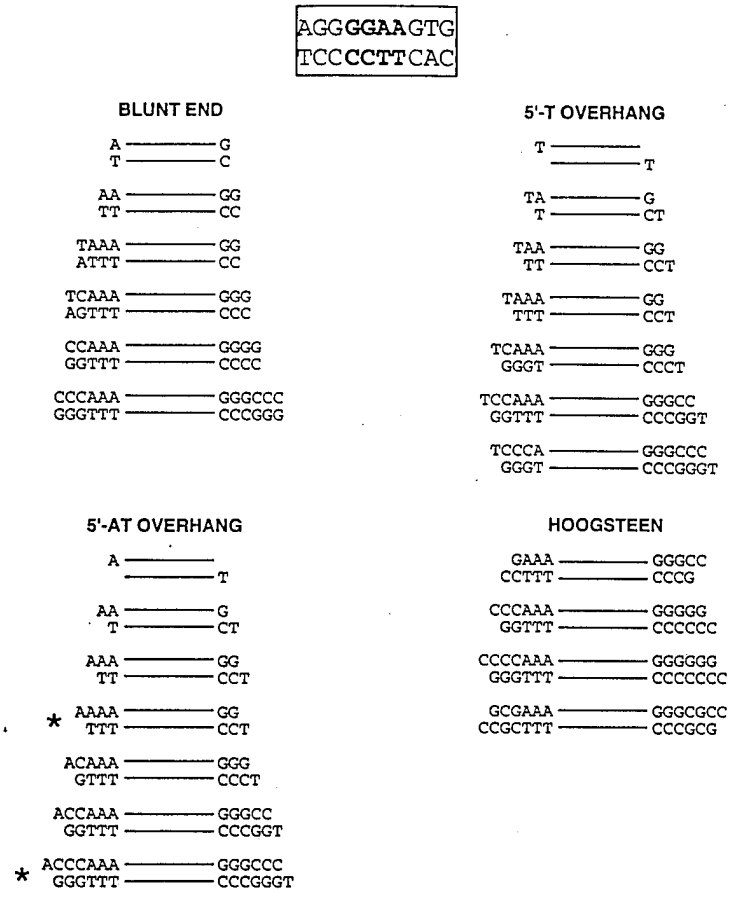


Figure 2. Oligonucleotides tested in co-crystallization trials. Each of the oligonucleotides listed were synthesized for co-crystallization with the PU.1 domain. The sequences differ in length and termini flanking a core sequence shown in the box at the top of the figure. The core sequence contains the GGAA recognition sequence for PU.1 (bold). In each oligonucleotide, the lines represent the repetition of this same core sequence. The best success with the production of large crystals was achieved with two oligonucleotides with a 5'-AT overhang (marked with asterisks).

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

Samples of the PU.1 domain prepared in Task 1 were not stable in solution over the long periods of time required for three-dimensional triple resonance experiments. To ensure that no trace of protease had co-purified with the PU.1 domain, we tested a different purification scheme that is based on the DNA-binding properties of the protein rather on physical properties alone. In this procedure, the protein was first fractionated on Affi-Gel Blue resin known to bind nucleotide-binding proteins. It was possible to achieve a remarkable level of purification at this step even with crude cell lysate. Next, the eluted fractions containing the PU.1 domain were applied to a hydroxyapatite column. This matrix is frequently used to isolate nucleic acid-binding proteins since it mimics the phosphate backbone recognized by such proteins. The PU.1 domain was eluted from the resin at pH 5.5 with 1M potassium phosphate buffer. It was also possible in this same step to concentrate the protein in an acidic buffer required for slow amide exchange in the NMR experiments. For both chromatographic steps, we were conservative in the selection of fractions that contained PU.1. The isolated protein was extremely pure and stable as judged by SDS-PAGE electrophoresis (see Figure 3).

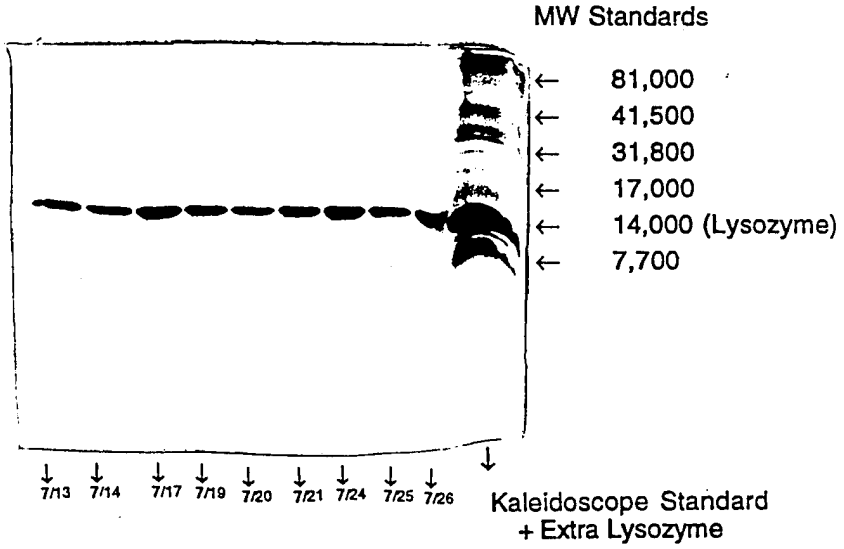


Figure 3. SDS-Page analysis of the stability of the PU.1 ETS domain. A sample of the PU.1 domain was placed in a 30°C water bath. Ailiquots were removed each day and frozen at -70 °C. At the end of 14 days, the aliquots were analyzed electrophoretically by SDS-PAGE. As can be seen from the electrophoretic pattern of selected aliquots, there is no degradation of the domain even after 14 days at 30°C (right lane). Molecular weight standards are shown for comparison in the right lane.

The purified protein was subjected to a stringent stability test for 14 days at 30°C, as shown in Figure 3. The same stability could be achieved with the introduction of the hydroxyapatite fractionation following ion-exchange chromatography, indicating that this is an essential step in preparation of PU.1 samples for NMR analysis. The stability of the domain after long term storage and data acquisition was tested by

MALDI mass spectroscopy. When an aliquot of a concentrated sample taken directly from the NMR tube was tested, the reported mass was 13,200 KD; the calculated molecular weight for the PU.1 domain is 13,089 KD.

Concentrated samples were prepared for NMR analysis. A systematic search was initiated to identify the most favorable combination of temperature, pH, ionic strength, protein concentration and buffer conditions needed to maintain proper protein conformation and avoid aggregation at the concentrations needed for NMR studies. Spectra taken in phosphate buffer, pH 5.5, with a trace of sodium azide, showed good linewidths and appropriate chemical shift dispersion. These conditions were used to prepare the less abundant isotopically labeled samples: a ^{15}N sample and a $^{13}C,^{15}N$ doubly labeled sample. The additional purification steps ensure the stability of the labeled samples, and these results represent a significant accomplishment for the success of the NMR experiments.

Subtask a. Heteronuclear resonance data will be collected from labeled samples at various pH values and temperatures.

An ^{15}N sample and a $^{13}C,^{15}N$ doubly-labeled sample were prepared from *E. coli* cultures grown in minimal media at room temperature and provided with $^{15}NH_4Cl$ and ^{13}C-glucose as the sole sources of carbon and nitrogen as described by Muchmore et al. (22). These samples were purified as described in the previous section and are stable in solution. Examples of two heteronuclear experiments are shown in the spectra in Figure 4.

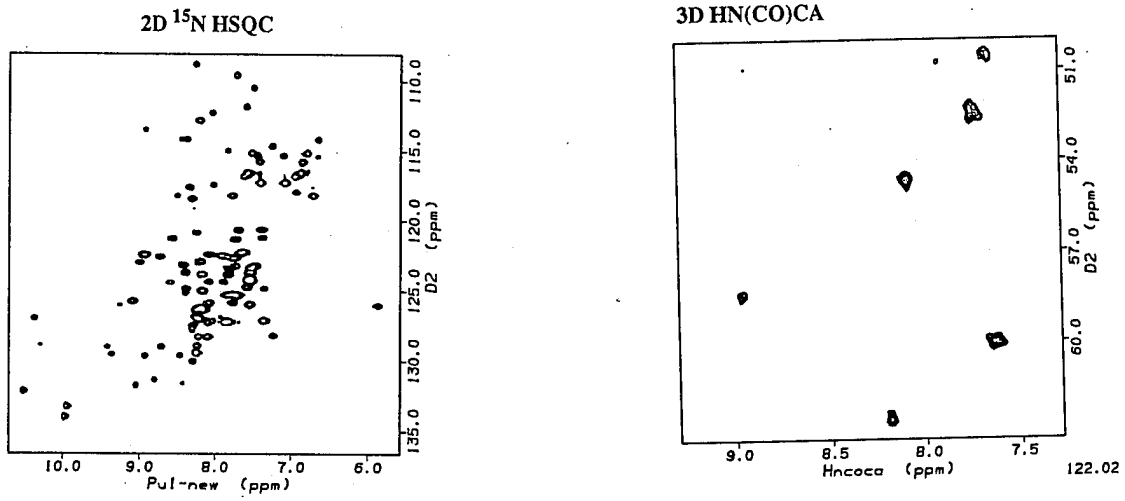


Figure 4: The 2D ^{15}N-^{1}H HSQC of the PU.1 DNA-binding domain. Most of the backbone NH amides as well as 3 indole-NH's from the side chains of 3 Trp residues are shown. Asn and Gln side chain NH_2 resonances are also observed. A slice corresponding to the ^{15}N plane at a resonance of 122 ppm from the 3D HN-CO-Ca experiment is also shown.

We have also performed a number of homonuclear experiments, all at 30 °C and a few test experiments at 24 °C. A partial list of the data acquired to date follows:

Homonuclear Experiments:

2 TOCSY experiments
1 Double Quantum experiment
1 2QF-COSY experiment

Heteronuclear Experiments:

^{15}N-HSQC
3D-^{15}N-NOESY-HSQC
3D-^{15}N-TOCSY-HSQC
^{13}C-HSQC
3D-HN-CO-CA
3D-HNCA

With data acquired from these experiments on stable samples the process of amino acid specific assignment is well underway; all the aromatic backbone resonances and partial side chain resonances have been asssigned and identification of all other resonances as well as sequential assignments are in progress. Though the HSQC spectra are quite good, on careful count there are approximately 15 backbone amide crosspeaks not readily apparent. Because of this discrepancy, and the parallel observation of fewer crosspeaks than expected in the 2-dimensional data, we have proceeded very cautiously on this stage of the work and spent a big effort verifying the integrity of our sample as outlined in the previous section. Since we have confirmed that the sample contains protein of the expected length, the source of this apparent lack of crosspeaks may reside in intrinsic properties of the DNA-binding domain itself. In the free state, the domain may not be entirely compact, and regions that are quite flexible may therefore contribute to "conformational averaging". Such plasticity of DNA-binding domains is not unprecedented; additional folding upon binding to DNA has been reported for the Trp repressor (22,23), leading to better defined secondary structure elements. Interestingly, the converse effect, unfolding upon DNA binding, has also been reported for the BAM H1 endonuclease-DNA complex (24). Experiments are now being planned to evaluate this conformational plasticity in the PU.1 domain.

Task 4: Determination of the crystal structure of the PU.1 domain complexed to DNA: Months 6-36

Subtask a. DNA oligonucleotides will be complexed to the PU.1 domain and tested for crystallization.

As a test of the protein samples, both of the protein fragments, differing in length, were tested for crystallization alone. The shorter fragment did not behave well in solution so it was not surprising that the fragment produced no crystals until it was mixed with DNA. However, it was surprising that the longer fragment failed to crystallize alone, since it was monodisperse in solution as measured by dynamic light scattering. This method measures the translational diffusion coefficient of a macromolecule. When performed in solution prior to crystallization, these analyses can be used to predict molecular samples that are not aggregated and likely to crystallize (26,27). As stated earlier, it is not unusual for DNA-binding proteins to crystallize only when complexed to DNA. And, in fact, our goal in this project was to determine the crystal structure of the PU.1 ETS domain-DNA complex, so we proceeded directly to experiments testing the formation of these complexes with the two protein fragments and the several oligonucleotides purified in Tasks 1 and 2.

Prior to mixing with protein, duplex DNA was annealed by heating to 95 °C and then slowly cooling to 20 °C. Duplex DNA oligos shown in Figure 2 were mixed with protein in molar ratios of 2:1 or 1:1 DNA:protein. The formation of complex was verified by gel shift electrophoretic assays. Solubility testing and precipitation testing were performed with selected complexes before crystallization trials. The solubility of the protein-DNA complexes was diminished relative to the protein alone. In fact, some of the complexes precipitated immediately upon mixing. These precipitates could be prevented if NaCl was present in the protein solution. The optimal concentration of NaCl differed for each complex.

PU.1-DNA complexes were formed with each of the oligonucleotides shown in Figure 2 and each of the two PU.1 fragments. These complexes were screened for crystallization using the sparse matrix method (28), starting with oligonucleotides > 20 bp in length. In these initial screens, crystals grew from conditions that are typical for protein-DNA complexes, i.e., neutral pH, polyethyleneglycol (PEG) and divalent cations (29). With these promising preliminary results, we moved on to the next task with efforts to increase the size and quality of the crystals.

Subtask b. Crystallization conditions will be modified and/or seeding methods will be implemented to produce large diffraction-quality crystals.

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. Conditions required to grow these crystals suggested that acetate was essential for crystallization. Indeed, further screening altering the pH and the acetate concentration produced larger crystals of the complex in two months.

Next, the shorter nucleotides shown in Figure 2 were tested. Those with the AT-overhang were given priority because of the results with the 23 bp oligo. From this

screening, we discovered that the long protein fragment complexed to a 16 bp oligonucleotide produced crystals readily. However, under the conditions describe for the 23 bp oligo complex, only crystals with an irregular morphology were obtain With further screening, well-shaped crystals grew in drops that contained PEG and zinc acetate. In the literature, a number of helix-turn-helix (HTH) DNA-binding prot have been crystallized from PEG solutions in acetate buffers, but to our knowledge is the first example of a HTH protein crystallized in the presence of zinc acetate. Tl observation that both the zinc and the acetate ions promote crystallization of ETS domain-DNA complexes may be of general utility for crystallization of other ets proteins. The zinc ion may stabilize the protein structure in the crystal, but confirma of this hypothesis awaits the elucidation of the crystal structure.

Final refinement of the crystallizaion conditions included altering the concentration and molecular weight of the PEG as precipitant. A dramatic improvement in crystal morphology was achieved by substituting PEG 600 for PEG 8000. Ultimately, larg crystals of the complex grew from solutions containing 100 mM cacodylate, pH6.5, 3-10% PEG 600 and 200 mM zinc acetate. Crystals formed in 3-5 days at 19°C. V have reported the crystallization of this complex, and a copy of the paper is include the APPENDIX:

Pio, F., Ni, C.Z., Mitchell, R.S., Knight, J., McKercher, S., Klemsz, M., Lombardo, A., Maki, R.A., 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., in press.

Subtask c. When large, high-quality crystals are obtained, high resolution x-ray diffraction data will be collected.

The crystals of the PU.1-DNA complex belong to the space group *C*2 with *a*=89.1, *b*=101.9, *c*=55.6 Å and β=112.2°. There are two complexes in the asymmetric unit. The crystals are very birefringent and diffract to at least 2.3-Å resolution. However, they are sensitive in the x-ray beam. Therefore, crystals are flash-frozen before diffraction experiments in cryoprotectant solutions of 8% PEG 600, and 30% MPD. After freezing, the crystals are extremely stable in the x-ray beam at -145 °C with n significant decay after 2.5 days of data collection. A native data set that is 98% complete has been collected at -145 °C to 2.3-Å resolution, and the data collectior statistics are presented in Table 1.

Two approaches are being used to obtain heavy atom substitutions for phase calculation. One of these methods is traditional soaking of heavy metal compounc into existing crystals, but the other approach involves the covalent modification of t protein and/or DNA. Data sets have been collected for several heavy atom soaks the mercurial compounds (e.g., PCMB) are the most promising candidates for the multiple isomorphous replacement method. In the other approach, where covalen modification of the protein or DNA components is being tested, a significant effort l been directed to the production of these modified molecules. Ultimately, the purification protocols from Tasks 1 and 2 were used or adapted for the modified molecules in order to produce milligram quantities of these "customized" molecule

In order to produce modified protein for MAD (multiwavelength anomalous dispersion) phasing methods (30), recombinant PU.1 domain was produced as a selenomethionine-substituted protein in *E. coli* B834 cells which are auxotrophic for methionine, using selenomethionine supplemented as the sole methionine source. The growth of these cells was slow, but the expression level was sufficient to produce milligrams of the modified protein. The presence of the selenomethionine was confirmed by amino acid analysis and the extent of substitution was shown to be 70-86%. Large crystals were produced with this protein complexed with DNA. X-ray data were collected from frozen crystals of this modified complex at multiple wavelengths at the LURE synchrotron source in Orsay, France. There are three methionines in the PU.1 domain, but the anomalous signal from these modified crystals was not sufficiently strong to be useful for phase calculation.

For the production of modified DNAs, we have substituted iodinated uracil phosphoramidites for thymine phosphoramidites in the synthesis of the oligonucleotides. Three iodinated oligonucleotides were synthesized with the iodine substituted at two sites on one strand and at a third site on the complementary strand. Large crystals have been obtained with each of these modified oligonucleotides complexed with PU.1, and data sets to high resolution have been collected from frozen crystals of each of these three complexes.

The data sets for native and heavy atom crystals are now being used for Patterson searches and phase calculation. We have also collected anomalous data with the heavy atom crystals for use in MIRAS phase determination. Several promising derivatives have been obtained. In addition, besides serving as sites for heavy atom substitution, the iodines may also serve as markers to orient the DNA in the crystal lattice. The diffraction pattern from the native crystals displayed strong reflections near 3.5 Å that indicated that the DNA oligos lie approximately along the *b* axis. This information will be very useful in the initial interpretation of the electron density maps. The next phase of the project will be to calculate phases and generate electron density maps of the complex.

CONCLUSIONS AND FUTURE WORK

The results for the work in the first twelve months of the project have demonstrated that the PU.1 domain is a suitable candidate for structural studies. The progress toward crystallographic analysis of the protein-DNA complex has been very successful. In the next twelve months, data from the heavy atom substitutions will be used to calculate phases and electron density maps will be prepared. These maps will interpreted to trace the polypeptide chain as well as the DNA backbone. Atomic models of both components will be built into the electron density maps interactively at graphics workstations.

In this project, we will continue to place a strong emphasis on the solution studies by NMR since this is the only study where there is an oppurtunity to examine an ets molecule in the complex in the crystal as well as in solution. Our observation that there may be an inherent flexibility in the domain has quite interesting biological implications. Transcription factors must accurately and precisely locate and bind to

rather short DNA sequences within the context of a vast human genome. It is therefore not surprising that it has been shown that some proteins bind to DNA and this binding is accompanied by a conformational adjustment. In experiments planned for the coming year, we will probe whether there is conformational averaging due to flexilbility in the PU.1 domain by two approaches: a) alter conditions of unlabelled samples and ^{15}N-labelled samples to test whether at higher concentration, higher ionic strength and lower temperatures we may detect spectroscopically weak amide resonance crosspeaks; b) prepare blunt-ended oligonucleotides representative of the DNA-binding site (see Figure 2) and probe whether in the presence of double-stranded DNA there is evidence of more ordered structural elements. These experiments will be designed first using results from preliminary studies with simple circular dichroism (CD) analyses. CD studies require significantly less material and are quite diagnostic of the existence of structural elements. Once the proper conditions for the DNA site and the DNA binding are worked out, we will probe by NMR whether, as is the case for Trp repressor (22,23), we can obtain more helical constraints for the helical elements and/or observe evidence of better defined β-strands with the bound complex. In later stages of the project, the study of the backbone dynamics of the free and bound protein, compared with structural details of the complex in the crystal, will provide valuable information about DNA contacts by ETS domains and the intrinsic plasticity of the DNA binding surfaces of these important molecules.

REFERENCES

1. Karim, F., Urness, L., Thummel, C., Klemsz, M., McKercher, S., Celada, A., Van Beveren, C., Maki, R., Gunther, C., and Nye, J. (1990) The ETS-domain: a new DNA-binding motif that recognizes a purine-rich core DNA sequence. *Genes Develop.* **4**, 1451-1453.

2. 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.

3. Moreau-Gachelin, F., Ray. D., Mattei, M.G., Tambourin, R., and Tavitian, A. (1989) The putative oncogene Spi-1: murine chromosomal localization and transcriptional activation in murine acute erythroleukemias. *Oncogene* **4**, 1449-1456.

4. 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.

5. 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.

6. 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.

7. 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.

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. Watson, D.K., McWilliams, M.J., Lapis, P., Lautenberger, J.A., Schweinfest, C.W., and Papas, T.S. (1988) Mammalian ets-1 and ets-2 genes encode highly conserved proteins. *Proc. Natl. Acad. Sci. USA* **82**, 7294-7298.

11. Reddy, E.S.P. and Rao, V.N. (1988) Structure, expression, and alternative splicing of the human c-ets-1 proto-oncogene. *Oncogene Res.* **3**, 239-246.

12. Reddy, E.S., Rao, V.N. and Papas, T.S. (1987) The erg gene: a human gene related to the ets oncogene. *Proc. Natl. Acad. Sci. USA* **84**, 6131-6135.

13. Rao, V.N., Huebner, K., Isobe, M., Ar-Rusdi, A., Croce, C.M. and Reddy E.S.P. (1989) Elk, tissue-specific ets-related genes on chromosomes X and 14 near translocation breakpoints. *Science* **244**, 66-70.
14. Wasylyk, C., Kerckaert, J.-P., Wasylyk, B. (1992) A novel modulator domain of Ets transcription factors. *Genes Develop* . **6**, 965-974.
15. Lim, F., Kraut, N., Frampton, J., and Graf, T. (1992) DNA binding by c-Ets-1, but not v-Ets, is repressed by an intramolecular mechanism. *EMBO J.* **11**, 643-652.
16. 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 Struct. Biol.* **1**, 871-876.
17. Liang H., Olejniczak, E.T., Mao, X., Nettesheim, D.G., Yu, L., Thompson, C.B., and Fesik, S.W. (1994) The secondary structure of the ets domain of human Fli-1 resembles that of the helix-turn-helix DNA-binding motif of Escherichia Coli catabolite gene activator. *Proc. Natl. Acad. Sci. USA* **91**, 11655-11659.
18. Donaldson, L.W., Petersen, J.M., Graves, B.J. and McIntosh, L.P. (1994) Secondary structure of the ETS domain places murine Ets-1 in the superfamily of winged helix-turn-helix DNA-binding proteins. *Biochemistry* **33**, 13509-13516.
19. 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.
20. 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.
21. Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1988) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60-89.
22. Muchmore, D.C., McIntosh, L.P., Russell, C.B., Anderson,D.E. and Dahlquist, F.W. (1990) 'Expression and nitrogen-15 labelling of proteins for proton and nitrogen-15 nuclear magnetic resonance'. *Methods in Enzymology* **177**, 44-73.
23. Zhao, D., Arrowsmith, C., Jia, X. and Jardetzky, O. (1993) Refinement of the solution structures of the *E. coli trp* holo- and aporepressor. *J. Mol. Biol.* **229**, 735-746.
24. Zhang, H., Zhao, D., Revinton, M., Lee, W., Jia, X., Arrowsmith, C. and Jardetsky, O. The solutions structure of *trp* repressor-operator DNA complex. *J. Mol. Biol.* **238**, 592-614.
25. Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1995) Structure of bam HI endonuclease bound to DNA: Partial folding and unfolding on DNA binding. *Science* **269**, 656-663.

26. Mikol, V., Hirsch, E. and Giege, R. (1990) Diagnostic of precipitant for biomacromolecule crystallization by quasi-elastic light-scattering. *J. Mol. Biol.* **213**, 187-195.
27. Ferre-D'Amare, A.R. and Burley, S.K. (1994) Use of dynamic light scattering to assess crystallizability of macromolecules and macromolecular assemblies. *Structure.* **2**, 357-359.
28. Jancarik, J. and Kim, S.-H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.* **24**, 409-411.
29. 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.
30. Hendrickson, W.A., (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51-58.

Appendix

Preprint Appended

Journal of Biological Chemistry
IN PRESS

Co-crystallization of an ETS domain (PU.1) in Complex with DNA: Engineering the Length of Both Protein and Oligonucleotide

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, CA 92037

#To whom correspondence should be addressed: La Jolla Cancer Research Foundation, 10901 North Torrey Pines Road, La Jolla, CA 92037.

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

Running Title: Crystals of PU.1 ETS Domain-DNA Complex

[1]The abbreviations used are: IPTG, isopropyl-1-thio-β-D-galactopyranoside; PMSF, phenylmethylsulfonylfluoride; MAD, multiple anomalous dispersion; TEAB, triethylammoniumbicarbonate; PEG, polyethyleneglycol; MPD, 2-methyl-2,4-pentanediol; MIR, multiple isomorphous replacement.

SUMMARY

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 β=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 PEG 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 member of the *ets* family.

INTRODUCTION

Transcription factors bind to target DNA sequences and regulate important metabolic functions such as cell growth, development and differentiation. The PU.1 transcription factor (1) is a member of the *ets* gene family, a recently discovered family of regulatory proteins. There are now more

than 25 members in this family that have been identified in various organisms from *Drosophila* to humans (reviewed in References 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 carboxy-terminal end of the molecule in PU.1 (Reference 1; see Figure 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, Klemsz and Maki, unpublished results). 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). However, it should be remembered that proteins that are members of the large helix-turn-helix

family differ in secondary structural features that affect the relative orientation of the critical helices. These differences influence the specificity of DNA recognition. Similarly, it is likely that important structural distinctions will exist among members of the *ets* family. Moreover, the mode of DNA contact within the *ets* family still must be elucidated. In the fli-1 structural studies, intermolecular NOEs between ^{13}C-labelled protein and unlabelled DNA indicated that 7 residues were within 4 Å of the DNA and the results suggested that helix α3 was the recognition helix. In order to precisely define 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). The sequence of PU.1 is identical to the oncogene Spi-1 (17). Spi-1 is activated in the erythroid leukemia induced by Spleen Focus Forming Virus (SFFV). Integration of SFFV upstream of the Spi-1/PU.1 gene results in over-expression of the Spi-1/PU.1 protein. This event is associated with the development of erythroid leukemia. The normal function of PU.1 is still being characterized but it is already clear that this transcription factor is a regulatory protein for differentiation of monocytes and macrophages and for B cell maturation (reviewed in Reference 2). The 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. Twenty 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 References 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 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 PCR 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 to the original clone. For bacterial expression, pET plasmid constructs were used to transform *E. coli* BL21(DE3)pLysS cells. A preculture of 50 ml LB medium (25) and ampicillin (100 mg/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 L of LB-ampicillin media. Cells were grown overnight at 26 °C in an aerated fermentor (Microferm, New Brunswick, NJ). The next morning, 2.5 L of LB-ampicillin buffered at pH 7.4 with sodium phosphate were added to the culture. After warming to 26 °C, expression of protein was induced with the addition of 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG[1]). After 4 hours, cells were harvested by centrifugation and stored as a paste at -70 °C.

Purification of PU.1 DNA-binding Domain--Cell pellets from one 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 phenylmethylsulfonylfluoride (PMSF)]. Cells were lysed on ice by sonication, cell debris was cleared by centrifugation at 17,000 rpm and 4 °C for 60 minutes 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 minutes on ice. The precipitate was removed by centrifugation at 15,000 rpm and 4 °C for 30 minutes. The supernatant solution was dialyzed at pH 7.5 against 20 mM Tris-HCl, 60 mM NaCl and 0.1 mM PMSF and then centrifuged again before application to CM-Sepharose Fast-Flow resin. The PU.1 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 (MAD) 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 hours, 5 ml of this preculture was used to inoculate one 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 OD_{600} reached 0.15 and

expression of recombinant protein was induced by the addition of 1 mM IPTG. After 16 hours, 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 a 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 on a Vydac C4 column at 56 °C using an acetonitrile gradient in 100 mM triethylammonium bicarbonate (TEAB) buffer (pH 8.5). Fractions containing the full-length oligonucleotide were pooled and acetonitrile was removed by dialysis against TEAB buffer. The oligonucleotides were desalted in 20% ethanol on Biogel 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 polyethyleneglycol (PEG) and methylpentanediol (MPD) 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 encoded the minimal DNA-binding domain. These fragments are shown in Figure 1. The two fragments differ in length at both the amino- and carboxyl-terminal ends of the sequence. The N-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 Figure 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 domain excluding a region at the end of the PEST region that is a conserved hydrophilic sequence (see Figure 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 only crystallize 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 (Klemsz and Maki, unpublished results). A number of oligonucleotides were chemically synthesized that each included the PU box sequence and differed in length. As shown in Figure 2, oligonucleotides with termini that provide blunt-ended or overhanging bases were tested for co-crystallization. Each oligonucleotide was mixed with the purified PU.1 domain in solutions suitable for crystallization trials and tested for complex formation by non-denaturing gel electrophoresis (results not shown).

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 Biogel 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 slowly cooling 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 Figure 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. 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 Figure 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, polyethyleneglycol (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 Figure 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 essential for crystallization. Tests altering the pH and acetate concentration produced larger crystals of the complex (0.2 x 0.1 x 0.05 mm) after two months. These results were the first indication that the acetate ion was important for crystallization.

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

possible that both the acetate and the zinc ions promote the formation of large crystals of the PU.1-DNA complex. 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; 35), NFκB-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). Members of other families of DNA-binding proteins do not crystallize as frequently from acetate solutions. 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 to 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 further improved 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 cover slip 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 cover slips were inverted and sealed over the reservoir. Parallelopiped crystals formed at 19 °C in 3 to 5 days. In some case, macroseeding (41) was used to produce large crystals. Crystals were washed free of mother liquor, dissolved and subjected to non-denaturing 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% MPD. 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 *C*2 with *a*=89.1, *b*=101.9, *c*=55.6 Å and β=111.2°. Assuming a molecular weight for the complex of 22,800 daltons, calculations of the cell dimensions were consistent with Vm (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 1. 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 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. To test if this level of substitution is adequate for phasing by MAD methods, 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 (MIR) methods (e.g. Ref. 44). 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 MIR 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. By testing several combinations of protein and DNA, the ideal complex for packing in the crystal lattice was identified. 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, the authors thank Drs. Craig Hauser and Nuria Assa-Munt for helpful suggestions for bacterial culture, and Mrs. Khanh Nguyen for advice on HPLC chromatography. This work was supported by grant #DMD17-94J-4439 from the Department of the Army (KRE) and by National Institutes of Health grant #AI20194 (RAM).

FIGURE LEGENDS

Figure 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 labelled (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.

Figure 2. Oligonucleotides tested in co-crystallization trials. Each of the oligonucleotides listed were 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 as well as 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 sizeable 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 Hoogstein base-pairing between DNA fragments within the crystal lattice. Although the PU.1 domain bound these DNA fragments, crystals were never obtained for complexes formed with these oligonucleotides.

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) *Inter. 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 Develop.* **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. USA* **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. USA* **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 Develop.* **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* **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. USA* **91**, 11655-11659

14. Donaldson, L. W., Petersen, J. M., Graves, B. J., and McIntosh, L. P. (1994) *Biochem.* **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 Develop.* **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 edit., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

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., 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. USA* **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. Cryst.* **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. Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993) *Nature* **364**, 412-420

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

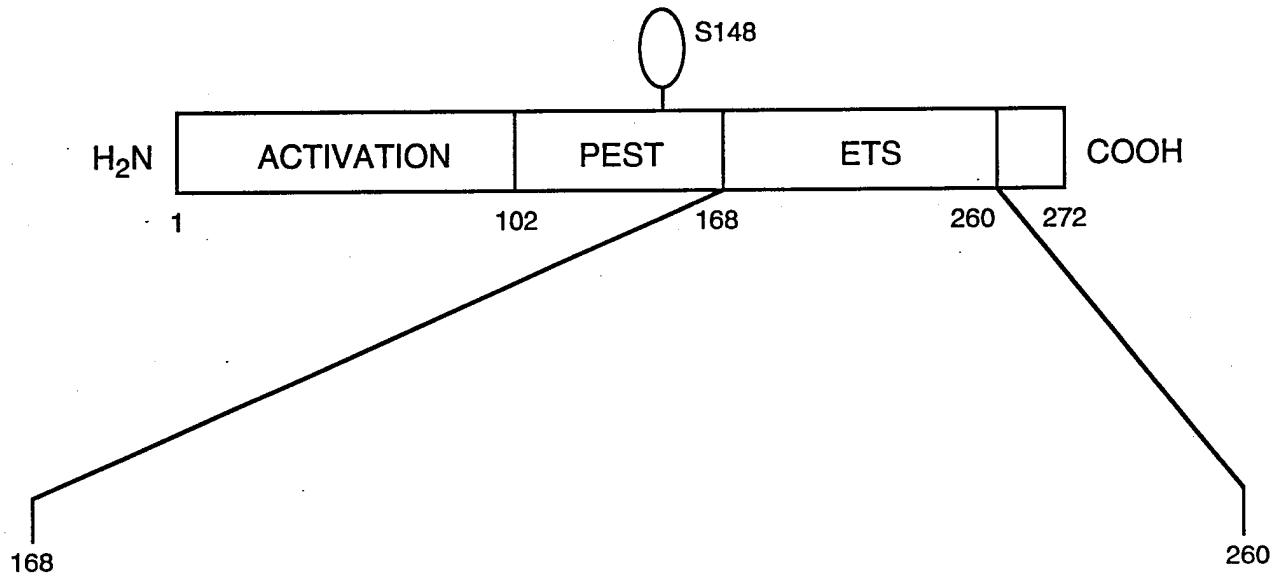
43. Low, B. W., and Richards, F. M. (1952) *J. Am. Chem. Soc.* **74**, 1660-1666

44. Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993) *Nature* **363**, 38-45

Table 1. Summary of data collection statistics

Minimum Resolution (Å)	Average Intensity (I)	Average I/σ(I)	Number of Observations	Number of Reflections	Rsym*
3.93	2898	48.3	17522	4063	0.040
3.12	2287	36.5	19299	4103	0.053
2.73	690	12.1	9339	4042	0.079
2.48	405	7.2	7256	3969	0.099
2.30	289	4.9	6679	3928	0.130
Totals	1327	22.0	60095	20105	0.050

* Rsym = $\Sigma | I_i - \langle I \rangle | / \Sigma I_i$, where I_i is the intensity of an individual measurement and $\langle I \rangle$ is the mean value of its equivalent reflections.



SKKKIRLYQFLLDLL.....TGEVKKVKKKLTYQFSGEVLG

SGLLHGETGSKKKIRLYQFLLDLL.....TGEVKKVKKKLTYQFSGEVLGRGGLAERRLPPH

ely illust a 7/95

AGG**GGAA**GTG
TCC**CCTT**CAC

BLUNT END

A ——— G
T ——— C

AA ——— GG
TT ——— CC

TAAA ——— GG
ATTT ——— CC

TCAAA ——— GGG
AGTTT ——— CCC

CCAAA ——— GGGG
GGTTT ——— CCCC

CCCAAA ——— GGGCCC
GGGTTT ——— CCCGGG

5'-T OVERHANG

T ———
——— T

TA ——— G
T ——— CT

TAA ——— GG
TT ——— CCT

TAAA ——— GG
TTT ——— CCT

TCAAA ——— GGG
GGGT ——— CCCT

TCCAAA ——— GGGCC
GGTTT ——— CCCGGT

TCCCA ——— GGGCCC
GGGT ——— CCCGGGT

5'-AT OVERHANG

A ———
——— T

AA ——— G
T ——— CT

AAA ——— GG
TT ——— CCT

* AAAA ——— GG
TTT ——— CCT

ACAAA ——— GGG
GTTT ——— CCCT

ACCAAA ——— GGGCC
GGTTT ——— CCCGGT

* ACCCAAA ——— GGGCCC
GGGTTT ——— CCCGGGT

HOOGSTEIN

GAAA ——— GGGCC
CCTTT ——— CCCG

CCCAAA ——— GGGGG
GGTTT ——— CCCCCC

CCCCAAA ——— GGGGGG
GGGTTT ——— CCCCCCC

GCGAAA ——— GGGCGCC
CCGCTTT ——— CCCGCG

ely illust b 7/95