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 C2 with a = 89.1 Å, b = 101.9 Å, 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 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). 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 Fil-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 Igκ gene in complex with a second factor NF-EMS (PIP) (18, 19). Formation of the ternary complex of PU.1, NF-EMS, 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 DAMD17-94-J-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, La Jolla, CA 92037. Tel.: 619-455-6480; Fax: 619-455-0181.

1 M. Klemsz and R. A. Maki, unpublished results.
Peripheral nervous system development, for example, relies on the presence of transcriptional repressors that bind to DNA sequences, such as the ETS family of proteins. The ETS family members are characterized by the presence of an ETS domain, a conserved motif consisting of 70 amino acids. The ETS domain contains a zinc finger-like structure that allows it to bind to DNA sequences containing a specific motif.
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 dialysis. 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.

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 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 mg/ml to 0.4 mg/ml. 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
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'-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 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% 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 β = 111.2°. Assuming a molecular mass for the complex of 22,800 daltons, calculations of the cell dimensions were consistent with Vm (42) of 2.58 Å³/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
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.

![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.

![Table I: Summary of data collection statistics](image)

| Minimum resolution (Å) | Average intensity | Average | Number of observations | Number of reflections | R~sym~ * |
|------------------------|-------------------|---------|------------------------|----------------------|---------|
| 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   |

* ~R~sym~ = \( \Sigma |I| - \langle I\rangle /\Sigma I\), where \( I\) is the intensity of an individual measurement and \( \langle I\rangle \) is the mean value of its equivalent reflections.
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. Ferre-D’Amare, 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. 1, 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