Crystal Structure of the Nuclear Matrix Targeting Signal of the Transcription Factor Acute Myelogenous Leukemia-1/Polyoma Enhancer-binding Protein 2αB/Core Binding Factor α2*

(Received for publication, June 7, 1999, and in revised form, July 23, 1999)

Lei Tang‡, Bo Guo§, Amjad Javed§, Je-Yong Choi§, Scott Hiebert‡, Jane B. Lian§, André J. van Wijnen§, Janet L. Stein§, Gary S. Stein§, and G. Wayne Zhou‡

From the ‡Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts 01605, the §Department of Cell Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01655, and the ¶Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, Tennessee 37232-0146

Transcription factors of the acute myelogenous leukemia (AML)/polyoma enhancer-binding protein (PEBP2α)/core-binding factor α (CBFA) class are key transactivators of tissue-specific genes of the hematopoietic and bone lineages. AML-1/PEBP2αB/CBFA2 proteins participating in transcription are associated with the nuclear matrix. This association is solely dependent on a highly conserved C-terminal protein segment, designated the nuclear matrix targeting signal (NMTS). The NMTS of AML-1 is physically distinct from the nuclear localization signal, operates autonomously, and supports transactivation. Our data indicate that the related AML-3 and AML-2 proteins are also targeted to the nuclear matrix in situ by analogous C-terminal domains. Here we report the first crystal structure of an NMTS in an AML-1 segment fused to glutathione S-transferase. The model of the NMTS consists of two loops connected by a flexible U-shaped peptide chain.

The spatial arrangement of the components of gene expression is important to their function. Chromosomal translocations that relocate the subnuclear targeting sequences of transcription factors can potentially contribute to the altered nuclear organization and related changes in gene expression that are characteristic of cancer cells. It is important, therefore, to understand how subnuclear targeting signals localize transcription factors.

The acute myelogenous leukemia (AML)1/core binding factor α (CBF-α/CBFA) transcription factors, as well as their homologs in mouse (polyoma enhancer-binding protein 2 (PEBP2) αA, αB, and αC) and Xenopus (Xaml), are key transactivators of tissue-specific genes of the hematopoietic and bone lineages (1–7). These CBFA/AML transcription factors provide a paradigm for studying tissue restricted transcriptional control within the context of nuclear organization. There are three distinct genes (AML-1/PEBP2αB/CBFA2, AML-2/PEBP2αA/CBFA3, and AML-3/PEBP2αC/CBFA1) that each encode both transcriptionally active and inactive isoforms which result from alternative mRNA splicing. The inactive forms lack the homologous C-terminal region.

The transcriptionally active form of the hematopoietic transcription factor AML-1 contains an unique C-terminal 31-amino acid nuclear matrix targeting signal (NMTS). This sequence directs AML-1 to nuclear matrix-associated subnuclear sites that support transcription. Furthermore, the NMTS of AML-1 can direct the heterologous GAL4 to the nuclear matrix (8, 9). Therefore, the NMTS represents an autonomous protein module with sufficient structural determinants to mediate subnuclear targeting. Domains analogous to the NMTS are conserved in the ubiquitous AML-2 and bone-related AML-3 regulatory proteins. Although subtle differences exist among the NMTS sequences of AML-1, AML-2, and AML-3, their NMTS segments may have similar targeting functions. The structure of the NMTS can, then, provide essential insight into properties of the AML factors that mediate trafficking to and/or binding to nuclear matrix-associated transcription sites. In this paper, we report the first crystal structure of the NMTS peptide, in this case fused with glutathione S-transferase, at 2.7-Å resolution.

**EXPERIMENTAL PROCEDURES**

**Crystallization and Data Collection—**Detailed procedures for protein expression, protein purification, crystallization, and data collection have been reported earlier (10). The space group of the crystal is P4₁2₁2₁, with cell dimensions of a = b = 93.4 Å, c = 57.6 Å. The final Rmerge of the data set is 11% and the data completeness is 94.9% up to 2.7 Å.

**Structure Determination and Refinement—**The phase determination of GST-NMTS was carried out by the molecular replacement method using AMoRe (11). The search model consisted of the coordinates of GST fused to a 6-amino acid peptide (Ref. 12, Protein Data Bank entry 1GNE), omitting the fused six residues, the water molecules, and glutathione. The molecular replacement was carried out with data between 10- and 4.0-Å resolution. Cross-rotation calculation produced a peak with a height of 9 o (3 o above the next highest non-symmetry related solution). The resulting solution from the translation search gave an R-factor of 35.5% and a correlation coefficient of 68.1%. After rigid body refinement of the molecular replacement solution using the resolution range of 8 Å-4 Å (implemented in AMoRe), the R-factor decreased to 33.6%. Simulated annealing was carried out with the program X-PLOR (13), followed by positioning and temperature factor refinement utilizing the restrained conjugate gradient least-square method. Refinement was conducted in the resolution range of 8 Å-2.7 Å. The starting R-factor was 38% and Rmerge was 38% (10% of the data were set off for free R-calculations). Residues 218–260 were built into 2Fo – Fc, σa, electron density map using the program O (14). No solvent molecules were added. Refinement continued until convergence was obtained. The final structure was refined to an Rmerge of 31.0% and an R-factor of 20.9%.

Refinement details and statistics are summarized in Table 1. The root mean square deviations of bond distances and bond angles from ideal geometry are 0.005 and 1.2 Å, respectively.

In *Situ* Nuclear Matrix Isolation and Indirect Immunofluorescence Analysis—*In situ* nuclear matrices were prepared as described (9). Ros...
between the Co atoms of the GST domain in the final model and the Co atoms of the Schistosoma japonicum GST structure is 0.9 Å. The link region forms an α-helix and a turn connecting the GST domain to the AML-1 segment. The NMTS region protrudes from the GST segment and resides in the space created by ordered GST moieties. As shown in Fig. 2A, the structure of the NMTS shows it can attain a finger-shaped loop region (Loop I), a hinge-shaped glycine-rich turn, and a β-strand (Strand II). These three regions are nearly coplanar. The approximate measured angle between the backbone of Loop I and the backbone of Strand II is 72.7°. The region of AML-1 adjacent to the NMTS (residues 233–237, represented in yellow in Fig. 1B) lies behind the plane formed by Loop I and Strand II. The η-atom of Arg-233 forms a weak hydrogen bond (3.3 Å) with the O-γ atom of Thr-241. There is also a 4.3-Å van der Waals contact between the side chains of Tyr-236 and Pro-244 (Fig. 1A; see also Fig. 4). These interactions may contribute to the stabilization of the NMTS conformation.

The Loop I Region of the NMTS Structure—Loop I, which contains several residues highly conserved among members of the AML family, is shaped like a finger. Although the side chain of Phe-240 is disordered, it should still bulge from the tip of the finger according to the position of its Cβ atom. The side chains of its adjacent residues, Thr-241 and Tyr-242, also protrude upwards from the tip (Fig. 2A). These three residues are highly conserved among AML-1 and AML-3 proteins (Fig. 3) and are, therefore, important in either nuclear matrix binding or in maintaining structural integrity. In the crystal structure, the tip region is followed by an Ser-Pro-Thr-Pro segment, which twists the position of the next two conserved residues, Val-247 and Thr-248, and forms a bump slightly above the plane of the NMTS segment. Ser-249, preserved among all members in the AML/PEBP2α/CBFA family, positions slightly beneath the plane and orients toward Strand II (Fig. 2A).

The Co atoms of the Loop I region exhibit high similarity to the CDR3 regions of the variable domains of antibodies (14). The two fragments which have the best matches are residues 94–103 of the Fab segment of immunoglobulin Kol (21) and residues 39–49 of the variable domain of the Bence-Jones protein REI (22). The fragments in Fab and REI belong to the antigen-binding site. The root mean square differences between the Co atoms are 1.6 and 1.8 Å for Fab and REI, respectively. This similarity suggests that the orientation of conserved amino acids in Loop I of the NMTS may have an analogous function in associating with a docking site in the nuclear matrix.

The Glycine-rich Turn of the NMTS Structure—A glycine-rich turn (GIGIG) is located at the bottom of Loop I and Strand II. The side chain of Ile-251 extends out from the bottom. The side chain of Ile-253 is disordered, although the position of its Cβ atom indicates that the alkyl side chain points to the inner face of the bottom of Loop I and Strand II. The Co atoms of the three glycine residues are all located on the outside surface of the turn (Fig. 2A). A similar glycine-rich motif has been associated with nucleotide-binding sites, such as the Rossmann fold (23), which contains a β strand followed by a glycine-rich loop and an α helix. In protein kinases and in protein phosphatases, the highly conserved P loop (GXGXXG) interacts with the phosphate group of the nucleotide. However, this glycine-rich sequence is not conserved among AML/PEBP2α/CBFA proteins (Fig. 3). Although the first glycine is preserved, the second glycine exists only in the AML-1 family and is replaced by serine in the AML-3 and AML-2 families. The last glycine is only conserved in the AML-1 and AML-3 proteins. A deletion occurs at this residue in AML-2. We propose that the GIGIG fragment may serve as a pivot point which permits rotation...
FIG. 1. A, the $2F_o - F_c$ electron density map around the C terminus of the GST-NMTS fusion protein, contoured at the 1.0 s level, was made using program O (14). The model, shown in white, consists of the link region (residues 218–232), the turn region (residues 233–237), and the NMTS...
between Loop I and Strand II to promote nuclear matrix association.

The Strand II Region of NMTS Structure—Strand II is composed of less conserved residues. Only the first three residues are conserved between AML-1 and AML-3: methionine, serine, and alanine. The remaining three residues Met-256, Gly-259, and Ser-260 are absent in the AML-3 family and substituted by serine, valine, and alanine in AML-2. Strand II is a β-strand with Ser-256, Met-258, and Ser-260 pointing toward Loop I (Fig. 2A). However, the side chain of Met-255 extends under the plane formed by Loop I and Strand II. It points away from the Loop I direction. The side chain of Ala-257 also faces away from Loop I.

Residues 261–274, which contain the C terminus of the NMTS region, are disordered in the final structure. As expected, the Ca atoms of Strand II are similar to those of the β-strands in other proteins. The closest fragment consists of residues 32–37 of REI, with an root mean square difference of 0.3 Å. In addition, the amino acids after residue 37 in REI form a loop in the crystal structure. Consistently, segments similar to Strand II in other proteins (e.g. acid proteinase, dihydrolate reductase, serine protease, and rhodanese) are all followed by surface loops. Furthermore, residues 378–383 of β-glycosidase (RYHLYM) (24), the segment in the RELIBase sequence data base with the highest sequence identity (67%) to the C-terminal region of the NMTS (RYHTYLY), forms a loop in the crystal structure. However, none of the entries in the sequence data base show similarity with residues 261 to 274 of NMTS. Hence, residues 261–262 in the NMTS are probably an extension of Strand II, while residues 263–268 adopt a loop conformation (Loop II, Fig. 2B). Therefore, we propose that Strand II and the C terminus of the NMTS comprise a second loop that participates in nuclear matrix association (Fig. 2B).

The Effect of Crystal Packing on the NMTS Structure—We have also considered the effects of crystal packing on the final NMTS structures. There is one hydrogen bond and one van der Waals interaction between residues in the NMTS and the symmetry-related GST (Fig. 4). The main chain carboxyl group of Gly-254 forms a 2.7-Å hydrogen bond with the Oδ-1 group of Asp-38 of the symmetry related molecule. In addition, Gly-259 has a 4.0-Å van der Waals' contact with Gln-43 of the neighboring molecule. These interactions may help to stabilize the β-strand in the crystal structure. Therefore, the structure of NMTS in solution may not be the same as the peptide structure reported here. However, we propose that the model we have presented, namely two finger-like loops connected by a U-shaped peptide chain, is a biologically meaningful representation of the NMTS peptide when it is a component of the native structure.
AML protein. We postulate that the two loops of the NMTS interact with a putative protein or nucleic acid acceptor in the nuclear matrix, as supported by mutagenesis results described below. The flexibility of the hinge region may accommodate the recognition.

Existing mutagenesis data are consistent with the proposed model. Deletion of 4 amino acids at the C terminus of the AML-3 NMTS almost completely eliminated nuclear matrix binding. In addition, a 3-amino acid mutation of the strand II region of the NMTS in mouse AML-1B (GMSA, residue 254–257, to KKSK mutation) did not affect nuclear matrix interaction.

The NMTS Sequence in AML-3 Is Necessary for Nuclear Matrix Targeting—The sequence alignment of AML proteins in Fig. 3 suggests that the conserved amino acids of the NMTS function in vivo to direct AML/PEBP2α/CBFA transcription factors to nuclear matrix-associated subnuclear sites that support transcription. The three-dimensional structure obtained from x-ray diffraction can then serve as a model for all nuclear matrix targeting signals in this family. We therefore examined whether AML-2 and AML-3 are nuclear matrix associated, as we have previously shown for AML-1. ROS 17/2.8 osteosarcoma cells were transfected with constructs expressing epitope-tagged AML-1, ROS 17/2.8 osteosarcoma cells were transfected with constructs expressing epitope-tagged AML-1, AML-2 (Fig. 5A), AML-2 (Fig. 5B), or AML-3 (Fig. 5C). Nuclear matrices were prepared, and the presence of AML factors was assessed by in situ immunofluorescence microscopy. The results shown in Fig. 5 indicate that AML-1, AML-2, and AML-3 each associate with the nuclear matrix.

We then addressed whether the NMTS of AML-3/CBFA1 resides in the C terminus, analogous to the location of the NMTS in AML-1. We generated a mutated AML-3/CBFA1 lacking the NMTS (C-terminal deletion of amino acids 362–513) and examined nuclear matrix targeting of this mutant protein by immunofluorescence microscopy. Fig. 6 illustrates that both the full-length AML-3 (Fig. 6B) and the C-terminal NMTS deletion (Fig. 6E) exhibit a punctate distribution in nuclei of detergent-extracted cells. DNA staining of the same cells is shown in Fig. 6, A and D, respectively. The full-length AML-3 is retained in the nuclear matrix (Fig. 6C), whereas the C-terminal deletion lacking the NMTS (Fig. 6F) is not.

---

* J.-Y. Choi, A. Javed, S. K. Zaidi, S. Hiebert, J. B. Lian, A. J. van Wijnen, J. L. Stein, and G. S. Stein, unpublished data.
nuclear import is retained but interaction with the nuclear matrix is abrogated. This result, together with our previous studies (26), establishes that the C termini of AML-3 and AML-1 each contain homologous intranuclear targeting signals.

Utilization of GST Fusion Proteins to Obtain the Crystal Structures of Small Peptides—The GST fusion protein expression system has been extensively used to generate large amounts of protein for structure studies. To prevent the interdomain flexibility introduced by the GST segment, GST fusion proteins are normally cleaved with proteases to release the GST moiety prior to crystallization. To date, the only published structure of a GST fusion protein is that of GST fused with a 6-amino acid conserved neutralizing epitope of gp41 from human immunodeficiency virus (12). We have determined the crystal structure of GST fused with a 41-amino acid peptide. Although the peptides attached to the C terminus of the GST protein were of different lengths, the crystal lattices are very similar between the two structures. Both belong to the tetragonal space group of P4 \(_2\) \(_2\) \(_2\). The unit cell dimensions differ by less than 1%. The fused peptides at the C terminus of the GST protein protrude from the GST domain and occupy the space formed by the crystal lattice of the GST protein. Therefore, the GST lattices may promote formation of the crystals. Taken together, these results support the concept that GST fusion proteins are useful protein vehicles to solve the crystal structures of small peptides.

Conclusions—Molecular recognition depends on steric or electrostatic complementarity of the combining surfaces. We report here that the structure of the NMTS of AML-1/PEBP2a/CBFA1 consists of two loops connected by a flexible U-shaped structure. The NMTS may interact with the nuclear matrix via the tips of the finger-like loop domains. We conclude that subnuclear targeting of CBFA/AML transcription factors to transcriptionally active sites is a highly selective process involving precise molecular recognition of specific components of nuclear architecture (8, 9, 27–29). The proposed model for the AML NMTS provides a basis for designing mutations that can further define structure-function interrelationships between the NMTS and the nuclear matrix.

Acknowledgment—We thank Elizabeth Bronstein for editorial assistance with the preparation of this manuscript.

REFERENCES
1. Miyoshi, H., Shimizu, K., Kezu, T., Maseki, N., Kaneko, Y., and Ohki, M. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 10431–10434
2. Ogawa, E., Maruyama, M., Kagoshima, H., Izuzuka, M., Lu, J., Satake, M., Shigemoto, K., and Ito, Y. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 6859–6863
3. Speck, N. A., and Stacy, T. (1995) Crit. Rev. Eukaryotic Gene Expression 5, 337–364
4. Meyers, S., and Hiebert, S. W. (1995) Crit. Rev. Eukaryotic Gene Expression 5,
365–383

5. Levano, D., Negreanu, V., Bernstein, Y., Bar-Am, I., Avivi, L., and Groner, Y. (1994) *Genomics* **23**, 425–432

6. Stein, G. S., van Wijnen, A. J., Stein, J. L., and Lian, J. B. (1998) *Front. Biosci.* **3**, D849–D864

7. Komori, T. (1998) *J. Bone Miner. Metabol.* **16**, 1–4

8. Zeng, C., van Wijnen, A. J., Stein, J. L., Meyers, S., Sun, W., Shopland, L., Lawrence, J. B., Penman, S., Lian, J. B., Stein, G. S., and Hiebert, S. W. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 6746–6751

9. Zeng, C., McNeil, S., Pockwinse, S., Nickerson, J. A., Shopland, L., Lawrence, J. B., Penman, S., Hiebert, S. W., Lian, J. B., Stein, A. J., Stein, J. L., and Stein, G. S. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 1585–1589

10. Tang, L., Guo, B., van Wijnen, A. J., Lian, J. B., Stein, J. L., Stein, G. S., and Zhou, G. W. (1998) *J. Struct. Biol.* **123**, 83–85

11. Navaza, J. (1994) *Acta Crystallogr. Sect.* **50**, 157–163

12. Lim, K., Ho, J. X., Keeling, K., Gilliland, G. L., Ji, X., Rucker, F., and Carter, D. C. (1994) *Protein Sci.* **3**, 2233–2244

13. Brünger, A. T. (1992) *X-PLOR Manual Version Version 3.1*, Yale University Press, New Haven, CT

14. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect.* **A47**, 110–119

15. Fey, E. G., Wan, K. M., and Penman, S. (1984) *J. Cell Biol.* **98**, 1973–1984

16. Reinemer, P., Dirr, H. W., Ladendstein, R., Huber, R., Lo Bello, M., Federici, G., and Parker, M. W. (1992) *J. Mol. Biol.* **227**, 214–226

17. Reinemer, P., Dirr, H. W., Ladendstein, R., Schaffer, J., Gallay, O., and Huber, R. (1991) *EMBO J.* **10**, 1997–2005

18. Ji, X., Zhang, P., Armstrong, R. N., and Gililand, G. L. (1992) *Biochemistry* **31**, 10169–10184

19. Sinning, I., Klewegt, G. J., Cowan, S. W., Reinemer, P., Dirr, H. W., Huber, R., Gililand, G. L., Armstrong, R. N., Ji, X., and Board, P. G. (1992) *J. Mol. Biol.* **232**, 192–212

20. Garcia-Saez, I., Parraga, A., Phillips, M. F., Mantle, T. J., and Coll, M. (1994) *J. Mol. Biol.* **237**, 298–314

21. Marguart, M., Deisenhofer, J., Huber, R., and Palm, W. (1980) *J. Mol. Biol.* **141**, 369–391

22. Epp, O., Laitman, E. E., Schiffer, M., Huber, R., and Palm, W. (1975) *Biochemistry* **14**, 4943–4952

23. Brahma, C., and Tooze, J. (1991) *Introduction to Protein Science*, Garland Publishing, New York

24. Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M., and Pearl, I. H. (1997) *J. Mol. Biol.* **271**, 789–802

25. Chen, L. F., Ito, K., Murakami, Y., and Ito, Y. (1998) *Mol. Cell. Biol.* **18**, 4165–4176

26. Lanford, R. E., Kanda, P., and Kennedy, R. C. (1986) *Cell* **46**, 575–582

27. Lamond, A. I., and Earnshaw, W. C. (1998) *Science* **280**, 547–553

28. Cook, P. (1998) *Science* **281**, 1466–1467

29. Wei, X., Samarabandu, J., Devdhar, R. S., Siegel, A. J., Acharya, R., and Berezney, R. (1998) *Science* **281**, 1502–1505

30. Evans, S. V. (1993) *J. Mol. Graph.* **11**, 134–138, 127–128