Identification and Characterization of SET, a Nuclear Phosphoprotein Encoded by the Translocation Break Point in Acute Undifferentiated Leukemia*

(Received for publication, July 16, 1993, and in revised form, October 4, 1993)

Yoshifumi Adachi, George N. Pavlakis, and Terry D. Copeland**

From the Human Retrovirus Section and the **Protein Structure Group, ABL-Basic Research Program, National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201

The translocation (6;9) in acute nonlymphocytic leukemia results in the formation of a dek-can fusion gene. In a case of acute undifferentiated leukemia, the oncogene can is fused to a different gene, named set, instead of dek and is assumed to be activated. Transcripts of set encode a putative SET protein with a predicted molecular mass of 32 kDa. We identified SET as a 39-kDa protein by immunoprecipitation with rabbit antisera against each of three synthetic peptides predicted from the open reading frame of the set gene. We confirmed this identification of SET by protein sequencing. We also observed that SET is expressed ubiquitously in various human cell lines. SET is phosphorylated on serine residue(s) in cultured cells and is localized predominantly in nuclei. Although the function(s) of SET and SET-CAN is not known, we propose that SET plays a key role in the mechanism of leukemogenesis in acute undifferentiated leukemia, perhaps by activating CAN in nuclei and stimulating the transformation potential of SET-CAN. This proposed role would therefore be similar to the roles observed for BCR and DEK of the chimeric oncoleukemia, perhaps by activating CAN in nuclei and acute nonlymphocytic leukemia, respectively.

The occurrence of defined chromosomal translocations in specific subtypes of leukemia strongly suggests that these translocations play an important role in the process of leukemogenesis. As a result of translocation, nearby oncogenes and cellular genes involved in the control of proliferation or differentiation can be activated through alterations in regulatory DNA sequences that leave the encoded protein intact (e.g. myc) (1) or through formation of fusion genes, which encode chimeric proteins (e.g. bcr-abl, E2A-pbx, and pre-BCR) (2-7).

The translocation (6;9) (p23;q34) in acute nonlymphocytic leukemia results in the formation of a highly consistent dek-can fusion gene (8). Translocation break points invariably take place in single introns of dek and can, named icb-6 and icb-9, respectively. In the case of acute undifferentiated leukemia (AUL), a break point was detected in icb-9 of can, whereas no break point was detected in dek. Genomic and cDNA cloning showed that instead of dek, a different gene was fused to can, which was named set (9). The set gene is located on chromosome 9q34, centromeric of c-abl. The set gene encodes transcripts of 2.0 and 2.7 kilobase pairs that result from the use of alternative polyadenylation sites. Both transcripts contain the open reading frame for a putative SET protein with a predicted molecular mass of 32 kDa (9). The set-can fusion gene in AUL encodes a 5-kilobase pair transcript that contains a single open reading frame predicting a 155-kDa chimeric SET-CAN protein. The set sequence shows homology to the yeast nucleosome assembly protein NAP-1 and shares the common sequence motif with DEK, B-23, HMG-1, and HMG-2 proteins (8, 10-12) at the COOH-terminal acidic region. SET has a long acidic tail, a large part of which is present in the predicted SET-CAN fusion protein. Despite rapid progress in molecular research on the set and set-can genes (9), little is known about their gene products, SET and SET-CAN.

Here, we report the identification of SET, a 39-kDa protein encoded by the set gene associated with the translocation break point of can in AUL. We show that (i) expression of SET is ubiquitous in various human cell lines; (ii) SET is phosphorylated in vivo, mainly on serine residues; and (iii) SET is located predominantly in nuclei. Our findings should be helpful in studying the function of SET and chimeric SET-CAN proteins in the mechanism of leukemogenesis in AUL.

EXPERIMENTAL PROCEDURES

Cells and Metabolic Labeling—HUT-102, MT-2, and TL-Su (13) are human T-cell lines infected with human T-cell leukemia virus type I. HUT-78 (14), I-9 (15) and Jurkat (16) are uninfected human T-cell lines. Raji (13) and Daudi (17) are human B-cell lines. Cell lines K-562 (13) and HL-60 (18) are human erythroleukemia cells and promyelocytic cells, respectively. Cell lines HOS, HeLa (19), and BIM (20) are human osteosarcoma cells, epitheloid carcinoma cells, and neuroblastoma cells, respectively. Labeling of cells with [35S]methionine or [32P]orthophosphate was conducted by incubating cells (1 x 10^6 cells/ml) in methionine- or phosphate-free medium supplemented with 10% fetal calf serum dialyzed against Tris-buffered saline (20 mM Tris-HCl, 0.15 M NaCl, pH 7.5) at 37 °C for 8 or 2 h, respectively (16, 21). The concentration of isotope added to the medium was 7.4 MBq/ml for [35S]methionine or 18.5 MBq/ml for [32P]orthophosphate.

Peptide Synthesis and Antibodies—Oligopeptides were synthesized with an automated peptide synthesizer (Applied Biosystems Model 430A) using t-butoxycarbonyl amino acids and p-methylbenzhydrylamine resin (16, 19, 22). Peptides were conjugated to keyhole limpet hemocyanin. Antiserum against each conjugated peptide was raised in rabbits as described previously (22). Immunoprecipitation and Electrophoresis—Metabolically labeled cells were lysed with radiolabeled precipitation buffer (60 mM Tris-HCl, 0.15 M NaCl, 4 mM EDTA, 1% Nonidet P-40, 0.1% sodium deoxycholate, 10 mM Na3P04, 10 mM NaF, 2 mM NaN3, 1 mM phenylmethylsulfonyl fluoride, pH 7.5) at 4 °C for 30 min. After centrifugation at 5000 x g for 30 min, supernatants were incubated with anti-SET serum in the absence or presence of peptide competitors at 4 °C for 2 h. The
immunocomplexes were precipitated with protein A-Sepharose CL-4B as previously described (16, 21). The immunoprecipitates were subjected to SDS-PAGE (23) followed by autoradiography or fluorography with Amplify (Amersham Corp.).

1. MSAQAAYKSKELNSNHDGADTSEKQEALENIDENVEI1DRNLEQAS 50
2. EELKEYQDNKLRQPRFKQSELIAKIPNWTVTTWNNIVQASALLGED 100
3. EEALHYTBEYTEFEDIKQ Y IDYDFEONPFINYLSKEHFLNESGD 150
4. PSSKTEIKWSDKGKLQSSOTOKASRKEQHPEFESFTWTDNHSAG 200
5. ADELEYKVDCQNPWPLQLYYLVPMDDEEEGEDZEDDDDEEEGEDIDE 250
6. EDEDEGEEGDGDEGEDGE277

Fig. 1. Predicted amino acid sequence of set gene product, SET (9), and synthetic peptides as shown: SP-1 (residues 3-16), SP-2 (residues 44-56), and SP-3 (residues 169-181). Dashed lines indicate the amino acid sequences of the synthetic peptides for antibodies used in this study. Boldface letters indicate highly acidic sequences. The position where SET is predicted to be fused to CAN in the set-can gene fusion in an AUL patient is indicated with an arrowhead (position 270) (9). Underlining indicates the consensus sequence of the sites phosphorylated by protein kinases (25, 26).

FIG. 2. Detection of SET protein. Human erythroleukemia K-562 cells were metabolically labeled with [35S]methionine for 8 h at 37 °C. Cell lysates were subjected to immunoprecipitation using antisera to synthetic peptides of SET (see Fig. 1) and to SDS-PAGE followed by fluorography. Lanes 1, 4, and 7, preimmune serum; lane 2, anti-SP-1 serum; lane 3, anti-SP-1 preabsorbed with SP-1; lane 4, anti-SP-2 serum; lane 5, anti-SP-2 preabsorbed with SP-2; lane 6, anti-SP-3 preabsorbed with SP-2; lane 8, anti-SP-3 serum; lane 9, anti-SP-3 preabsorbed with SP-3. Molecular size standards are indicated on the left.

Fig. 3. Identification of SET protein. A, isolation of SET from K-562 cells by large-scale immunoprecipitation. Cell extracts from 1 × 10⁶ cells were immunoprecipitated with preimmune serum (lane 1) and anti-SP-1 serum (lane 2). The immunocomplexes were separated by SDS-PAGE and detected by Coomassie Brilliant Blue R-250 staining. Arrows indicate separated immunoglobulin heavy (IgG(H)) and light (IgG(L)) chains and SET. Lane M contains molecular size standards. B, partial amino acid sequence of SET. Isolated SET from gels was subjected to TLCK-treated chymotrypsin digestion. Microsequencing yielded the two peptide sequences shown in the shaded boxes and aligned with the predicted sequence.

Protein Sequencing—The 39-kDa protein was digested with TLCK-treated chymotrypsin. After reverse-phase HPLC using a C₁₈ column (2 × 250 mm; Vydac), the separated peptides were subjected to automated Edman degradation on a pulse liquid Sequencer (Applied Biosystems Model 477A). Phenylthiohydantoin-derivatives were identified with an on-line Model 120 phenylthiohydantoin-derivative analyzer (19).

Identification and Characterization of SET

RESULTS

Identification of SET Protein—The nucleotide sequence of the set cDNA contains an open reading frame encoding a protein of 277 amino acids with a predicted molecular mass of 32 kDa, as shown in Fig. 1 (9). To detect the SET protein, we prepared three rabbit antisera, each directed against a synthetic peptide from a different portion of the putative SET protein. Human erythroleukemia K-562 cells were metabolically labeled with [35S]methionine for immunoprecipitation. A predominant 39-kDa protein was immunoprecipitated from these cells with antisera directed against the SP-1 (residues 3-16), SP-2 (residues 44-56), and SP-3 (residues 169-181) peptides, respectively (Fig. 2, lanes 2, 5, and 8), but not with preimmune serum (lanes 1, 4, and 7). A 66-kDa protein was also detected by immunoprecipitation with anti-SP-1 and anti-SP-2 sera, but the intensity of this 66-kDa protein was less than one-twentieth of the major 39-kDa protein. The immunoprecipitation of these proteins could be completely prevented by preabsorption with the same peptides used for immunization (lanes 3, 6, and 9), indicating specific antigen-antibody reactions. Thus, the major 39-kDa protein contained three sequences of the putative SET protein (residues 3-16, 44-56, and 169-181).

To obtain more information for the identification of the 39-kDa protein, especially for protein sequence, we isolated the 39-kDa protein from K-562 cells by large-scale immunoprecipitation using anti-SP-1 serum and preparative SDS-PAGE (Fig.
Identification and Characterization of SET

**FIG. 4. Expression of SET protein in various human cell lines.** Cell extracts (60 μg of protein in every case) from various human cell lines were subjected to immunoblot analysis using anti-SP-1 serum. Lane M, prestained molecular size standards; lane 1, HUT-102; lane 2, MT-2; lane 3, TL-Su; lane 4, HUT-78; lane 5, H-9; lane 6, Jurkat; lane 7, Raji; lane 8, Daudi; lane 9, K-562; lane 10, HL-60; lane 11, HeLa; lane 12, HOS; lane 13, BIM.

3A, (lane 2). Bands migrating with an apparent molecular mass of 39 kDa were excised and subjected to TLCK-treated chymotrypsin digestion, reverse-phase HPLC separation, and microsequencing. The sequence obtained for peptide 1 (with recoveries in picomoles in parentheses) was Ile(8)-Ala(7)-Lys(2)-Ile(2)-Lys(1)-Asp(1). A data base search revealed that both sequences are identical to the predicted sequences from set cDNA (9). In particular, the first peptide sequence corresponds to residues 191-210 (Fig. 3B). These sequences are identical to the predicted sequences from set cDNA (9).

**Expression of SET in Various Human Cell Lines—**To test whether SET protein is expressed ubiquitously in other cells, we subjected cell extracts (60 μg of protein) from various human cell lines (HUT-102, MT-2, TL-Su, HUT-78, H-9, Jurkat, Raji, Daudi, K-562, HL-60, HeLa, HOS, and BIM) to immunoblot analysis using anti-SP-1 serum. As shown in Fig. 4, all of the cell lines that we employed expressed SET at approximately the same level. The apparent molecular mass of SET in the immunobLOTS was the same, 39 kDa. Thus, the expression of SET at the protein level was ubiquitous.

**In Vivo Phosphorylation of SET—**Since the amino acid sequence of SET contains the apparent consensus sequences of the sites phosphorylated by protein kinases (Fig. 1) (25, 26), we attempted to determine whether in vivo phosphorylation of SET occurred in cells. K-562 cells were metabolically labeled with [32P]orthophosphate and subjected to immunoprecipitation. Radioactive SET was then immunoprecipitated with antisp-1 serum (Fig. 5A, lane 2). Because the radioactivity associated with SET was sensitive to 50 μg/ml alkaline phosphatase treatment at 37 °C for 2 h (lane 3), the incorporation of radioactivity was due to protein phosphorylation, but not ADP-ribosylation.

To determine the phosphoamino acid content of SET, [32P]labeled SET from K-562 cells (shown in Fig. 5A, lane 2) was eluted from the gel and hydrolyzed in 6 N HCl at 110 °C for 2 h. Phosphoamino acids were separated by electrophoresis and autoradiographed. As shown in Fig. 5B, SET protein was phosphorylated mainly on the serine residue in vivo, whereas phosphothreonine and phosphotyrosine were not detected.

**Subcellular Localization of SET—**We investigated the subcellular localization of SET in HeLa and HOS cells by indirect immunofluorescent staining with anti-SP-1 serum. Fig. 6A shows typical HeLa cells stained using anti-SP-1 serum. The immunofluorescent signal was observed mainly in nuclei. The outer layer of the nuclear membrane was also stained weakly.

**DISCUSSION**

In this report, we identified and characterized a 39-kDa protein as SET protein, which was predicted from the open reading frame of set transcripts (9). The criteria used to identify this protein are as follows. (i) The same 39-kDa protein could be immunoprecipitated with antisera to each of three peptides corresponding to different regions of the predicted amino acid sequence of SET (residues 3-16, 44-56, and 168-181) (Fig. 2), thus greatly decreasing the possibility that the immunoprecipitation was due to chance sequence homology between the syn-
Identification and Characterization of SET 2261

SET protein. HeLa (A and E) and HOS (B and F) cells were stained with anti-SP-1 serum and fluorescein isothiocyanate-labeled second antibody. The same cells were viewed by phase-contrast microscopy (C and D, respectively). Magnification x ~800 in A and C, x 540 in B and D, and x 400 in E and F.

Fig. 6. Subcellular localization of SET protein. HeLa (A and E) and HOS (B and F) cells were stained with anti-SP-1 serum and fluorescein isothiocyanate-labeled second antibody. The same cells were viewed by phase-contrast microscopy (C and D, respectively). Magnification x ~800 in A and C, x 540 in B and D, and x 400 in E and F.

The SET protein consists of 277 amino acids with a predicted molecular mass of 32 kDa (9). The discrepancy in the molecular size of the observed 39-kDa SET and the predicted 32-kDa SET may be due to protein phosphorylation (Fig. 5), the blocked NH₂ terminus, protein glycosylation, and/or the high content of acidic residues in the amino acid sequence (Fig. 1) (9) since phosphoproteins with blocked NH₂-terminal residues and a high content of acidic residues (e.g. nucleolar shuttle protein B-23, HMG-1, and HMG-2) have been shown to migrate more slowly than expected on SDS-PAGE (10-12, 19, 27).

SET is widely distributed in various human cell lines, as shown in Fig. 4. The molecular mass of SET observed on the immunoblot was the same (39 kDa) in every case. These results are consistent with previous observations of set gene expression at the mRNA level in mouse tissues and early embryos (9), suggesting that SET has a rather general function(s) in the cell.

The data described in this study clarify that SET is phosphorylated in vivo (Fig. 5). The major phosphoamino acid of phosphorylated SET was phosphoserine, indicating that SET is a substrate for one of the cellular serine/threonine kinases. Tyrosine kinases are not involved as phosphotyrosine was not detected. The function(s) of SET is not yet known; however, the in vivo phosphorylation of SET may be involved in the regulation of SET and SET-CAN functions or in the next event in the signal transduction pathway in response to physiological stimuli or the cell cycle.

SET was found predominantly in nuclei, as shown in Fig. 6. The nuclear localization of SET seems consistent as it contains an extremely high percentage of acidic residues, 32% (98 amino acids), half of which (43 amino acids) are present at the COOH terminus, forming a long acidic tail. Many proteins containing acidic regions are located in the nucleus and have different functions (28). Analogous to acidic domains in NAP-1, HMG-1, HMG-2, nucleolin, GAL4, and VP16 (12, 27, 29-33), the acidic motif of SET might serve as a nucleosome/chromatin assembly domain or a transcription activation domain.

Biological and functional assays are needed to determine whether the acidic domain of SET is essential for the putative transformation potential of the SET-CAN fusion protein in AUL. The nuclear localization of SET and its in vivo phospho-

---

2 Y. Adachi, T. D. Copeland, D. R. Kaplan, and G. N. Pavlakis, unpublished results.
rulation are interesting since fusion of CAN to DEK and of ABL to BCR results in a nuclear localization of the fusion proteins, whereas CAN and ABL themselves are present mainly in the cytoplasm (3, 8, 9, 34). Hence, fusion of SET to CAN may have the same effect, resulting in a nuclear localization of the SET-CAN fusion protein. A nuclear translocation of CAN may be essential for the putative leukemogenic effect of the fusion protein in AUL.

Acknowledgments—We thank Pat Wedock for peptide synthesis and purification, Suzanne Specht for preparation of antibodies, and Catherine Hixson for amino acid analysis. We are also grateful to Drs. Xian-kui Zhang and Mark Smith for helpful discussions and Anne Arthur for editorial assistance.

REFERENCES
1. Rabbits, T. H., Hamlyn, P. H., and Baer, R. (1983) Nature 306, 760-765
2. Shitileman, E., Lifshitz, B., Gale, R. P., and Canziani, E. (1986) Nature 315, 550-554
3. Hertman, A., Heisterkamp, N., von Lindern, M., van Baal, S., Meijer, D., van der Plas, D., Wiedeman, L. M., Groffen, J., Bootema, D., and Grosveld, G. (1987) Cell 61, 33-40
4. Kamps, M. P., Murre, M., and Balkens, D. (1989) Cell 60, 547-555
5. Nourse, J., Mullen, J. R., Galli, N., Wilkinson, J., Stanbridge, E., Smith, S. D., and Cleary, M. L. (1986) Cell 60, 535-546
6. De The, H., Lavau, C., Marchio, A., Chomienne, C., Degos, L., and Dejean, A. (1991) Cell 64, 673-684
7. Kikizuka, A., Miller, W. J., Umesono, K., Warrell, R. P., Frankel, S. R., Murty, V. V. S., Dmitrovsky, E., and Evans, R. M. (1991) Cell 64, 683-674
8. von Lindern, M., Forrester, M., van Baal, S., Jaegle, M., de Wit, T., Buja, A., and Grosveld, G. (1992) Mol. Cell. Biol. 12, 1687-1697
9. von Lindern, M., van Baal, S., Wiebert, J., Raap, A., Hageman, A., and Grosveld, G. (1992) Mol. Cell. Biol. 12, 3346-3355
10. Zhang, X. T., Thomas, D. C., and Samuel, C. E. (1989) Biochem. Biophys. Res. Commun. 164, 176-184
11. Wen, L., Huang, J. K., Johnson, B. H., and Rebeck, G. R. (1989) Nucleic Acids Res. 17, 1197-1214
12. Shirakawa, H., and Yoshida, M. (1992) J. Biol. Chem. 267, 6641-6645
13. Adachi, Y., Kitahara-Gawas, A., Sugenura, K., Lee, W. J., Yodoi, J., Maki, M., Murachi, T., and Hatanaka, M. (1992) J. Biol. Chem. 267, 19375-19378
14. Gazdar, A. F., Corney, D. N., Bunn, P. A., Russell, E. S., Jaffe, E. S., Schechter, G. P., and Gouros, J. G. (1980) Blood 55, 459-477
15. Popovic, M., Samusghadharan, M. G., Read, E., and Gallo, R. C. (1984) Science 224, 497-500
16. Adachi, Y., Copeland, T. D., Takahashi, C., Nosaka, T., Ahmed, A., Oroszlan, S., and Hatanaka, M. (1992) J. Biol. Chem. 267, 21977-21981
17. Klein, K., and Klein, G. (1987) Cancer Res. 36, 1300-1310
18. Adachi, Y., Takao, E., Murachi, T., and Hatanaka, M. (1988) Biol. Chem. Hoppe-Seyler 369, 223-227
19. Adachi, Y., Copeland, T. D., Hatanaka, M., and Oroszlan, S. (1993) J. Biol. Chem. 268, 13933-13934
20. Ishii, K., Adachi, Y., Hatanaka, M., Sakamoto, H., and Furuya, Y. (1990) J. Cell. Physiol. 143, 569-576
21. Adachi, Y., Ishida-Takahashi, A., Takahashi, C., Takanashi, M., and Hatanaka, M. (1991) J. Biol. Chem. 266, 3965-3972
22. Copeland, T. D., Tsai, W. P., Kim, Y. D., and Oroszlan, S. (1986) J. Immunol. 137, 2945-2951
23. Laemmli, U. K. (1970) Nature 227, 680-685
24. Noeske, T., Sjömi, M., Adachi, Y., Ishibashi, M., Kubota, S., Maki, M., and Hatanaka, M. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 9799-9802
25. Kishimoto, A., Nishiyama, K., Nakamichi, H., Utsatani, Y., Nomura, H., Takeyama, Y., and Nishiuaka, Y. (1985) J. Biol. Chem. 260, 12492-12499
26. Colbran, R. J., Scherer, C. M., Hashimoto, Y., Feng, Y. L., Rich, D. C., Smith, M. X., and Soderling, T. (1986) Biochem. J. 235, 513-525
27. Adachi, Y., Mizuno, S., and Yoshida, M. (1990) J. Chromatogr. 500, 39-46
28. Earnshaw, W. C. (1987) J. Cell Biol. 108, 1479-1482
29. Ishimi, I., and Kikuchi, A. (1991) J. Biol. Chem. 266, 7525-7529
30. Krohne, G., and Franke, W. W. (1988) Exp. Cell Res. 176, 167-189
31. Lapeyre, B., Bourbon, H., and Amadrid, F. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 1472-1477
32. Hope, I. A., and Struhl, K. (1986) Cell 65, 885-894
33. Brent, R., and Ptashne, M. (1985) Cell 43, 723-726
34. Poelinger, A. M., Muller, A. J., Havik, M. H., Maru, Y., and Witte, O. N. (1991) Cell 66, 161-171