

Nuclear punctate distribution of ALL-1 is conferred by distinct elements at the N terminus of the protein

(speckles/trithorax/11q23 abnormalities)

TAKAHIRO YANO*, TATSUYA NAKAMURA*†, JANNA BLECHMAN†‡, CLAUDIO SORIO*, CHI V. DANG§, BENJAMIN GEIGER‡, AND ELI CANAANI‡¶

*Kimmel Cancer Institute, Thomas Jefferson University, Philadelphia, PA 19107; †Department of Medicine, Cell Biology, and Anatomy, and The Johns Hopkins Oncology Center, The Johns Hopkins University School of Medicine, Baltimore, MD 21205; ‡Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel

Communicated by Carlo M. Croce, Thomas Jefferson University, Philadelphia, PA, April 28, 1997 (received for review March 27, 1997)

ABSTRACT The ALL-1 gene positioned at 11q23 is directly involved in human acute leukemia either through a variety of chromosome translocations or by partial tandem duplications. ALL-1 is the human homologue of *Drosophila trithorax* which plays a critical role in maintaining proper spatial and temporal expression of the *Antennapedia-bithorax* homeotic genes determining the fruit fly's body pattern. Utilizing specific antibodies, we found that the ALL-1 protein distributes in cultured cells in a nuclear punctate pattern. Several chimeric ALL-1 proteins encoded by products of the chromosome translocations and expressed in transfected cells showed similar speckles. Dissection of the ALL-1 protein identified within its $\approx 1,100$ N-terminal residues three polypeptides directing nuclear localization and at least two main domains conferring distribution in dots. The latter spanned two short sequences conserved with TRITHORAX. Enforced nuclear expression of other domains of ALL-1, such as the PHD (zinc) fingers and the SET motif, resulted in uniform nonpunctate patterns. This indicates that positioning of the ALL-1 protein in subnuclear structures is mediated via interactions of ALL-1 N-terminal elements. We suggest that the speckles represent protein complexes which contain multiple copies of the ALL-1 protein and are positioned at ALL-1 target sites on the chromatin. Therefore, the role of the N-terminal portion of ALL-1 is to direct the protein to its target genes.

The ALL-1 gene located at 11q23 was cloned (1, 2) because of its involvement in chromosome abnormalities occurring in 10% of patients with acute lymphocytic or myeloid leukemia, in particular in infants and secondary leukemias (3). In some of the patients, ALL-1 is undergoing partial tandem duplication (4), but in the majority of cases, ALL-1 recombines to a partner gene to produce a chimeric protein (1, 2). The number of partner genes is strikingly high—more than 25 are predicted by cytogenetic evidence, and 10 were already cloned (reviewed in refs. 5 and 6). The critical product of the translocations is composed of the N-terminal 1,300–1,400 residues of ALL-1 fused in phase to the C-terminal segment of the open reading frame of the partner genes. One such fusion product was shown to induce acute leukemia in mice (7).

On the basis of sequence homology, it was originally suggested (1, 2, 8) that ALL-1 is the human homologue of *Drosophila trithorax* (*trx*). Subsequent studies involving generation of ALL+/- knockout mice and of ALL-1-/- embryos, and the identification of homeotic transformations in these animals (9) proved this suggestion. In *Drosophila*, the activity

of the homeotic genes of the *Antennapedia* and *bithorax* complexes determine the various body structures along the anterior–posterior axis (10). Following the establishment of expression pattern of these homeotic genes shortly after fertilization by the segmentation genes, the state of determination has to be maintained during proliferation. The maintenance function is accomplished mainly by the genes of the *trithorax* and *Polycomb* groups (*trx-G* and *Pc-G*), named after their best known members (for reviews, see refs. 11 and 12). Genetic and molecular data indicate that the *trx-G* and *Pc-G* proteins act to trigger or inhibit transcription of their target genes, respectively. Applying histochemical analysis, it was found that the proteins encoded by several *Pc-G* genes show almost identical binding patterns to up to 100 genes on polytene chromosomes of salivary glands (13–15). The two *trx-G* proteins, TRITHORAX (TRX) and ASH1, also bind to polytene chromosomes at about 60 and 100 sites, respectively (16–18). Many of these sites colocalize with those of the *Pc-G* proteins. This suggests that *trx-G* and *Pc-G* proteins assemble on the chromatin into activating or repressive transcriptional complexes, respectively. Finally, it is thought that *trx-G/Pc-G* proteins act by remodeling chromatin. This is based on molecular similarities, as well as on physiological properties shared by some *trx-G/Pc-G* proteins with chromatin proteins that act as modifiers of position effect variegation. Direct biochemical evidence for chromatin alterations was recently obtained for the yeast/mammalian homologues of *Drosophila* BRM–SWI2/SNF2. These proteins are included within the SWI/SNF complex, which acts to disrupt nucleosomes and facilitates binding of transcription factors (19, 20). Similarly, the *trx-G* protein TRL (GAGA) is a component of another protein complex which remodels chromatin (21).

Two domains, the SET motif and PHD fingers, are highly conserved between the ALL-1 and TRX proteins. These domains are also found in some *trx-G/Pc-G* and related proteins, involved in embryonic development, and associated with chromatin (14, 17, 22, 23). ALL-1 and TRX share several additional short homologous regions (24–26). Finally, ALL-1 contains several unique motifs such as AT-hooks that are thought to bind DNA, transcriptional transactivation domain (TAD), and regions with homology to DNA methyltransferase (mT) or U1 small nuclear ribonucleoprotein (snRNP) (5, 6, 27). Here, we show that specific domains within the N-terminal portion of ALL-1 mediate speckled nuclear distribution of the protein. The possible meaning of these speckles is discussed.

Abbreviations: TRX, *Drosophila trithorax*; Ab, antibody; PK, pyruvate kinase; NTS, nuclear targeting sequence; SNL, speckled nuclear localization; mT, methyltransferase; snRNP, small nuclear ribonucleoprotein; TAD, transactivation domain.

†T.N. and J.B. contributed equally to this paper.

¶To whom reprint requests should be addressed. e-mail: licanani@dapsas1.weizmann.ac.il.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1997 by The National Academy of Sciences 0027-8424/97/947286-6\$2.00/0 PNAS is available online at <http://www.pnas.org>.

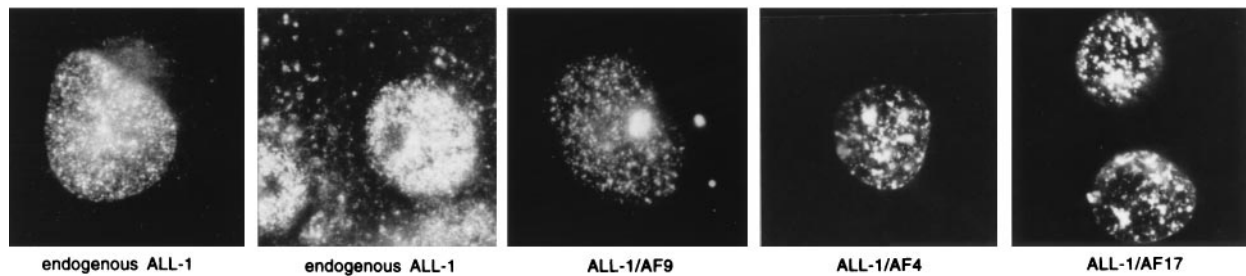


FIG. 1. Subcellular distribution in COS cells of endogenous ALL-1 and of transiently transfected chimeric ALL-1 proteins. FLAG-tagged ALL-1/AF-9 was detected with anti-FLAG mAb. The other proteins were detected with polyclonal Ab P3 or P4 against ALL-1. Transfected cells expressed at least 10-fold more exogenous protein compared with endogenous ALL-1.

MATERIALS AND METHODS

Plasmid Construction. Various restriction fragments of ALL-1 cDNA were subcloned into pSG5 expression vector (Stratagene) containing a FLAG-encoding sequence followed by a stop codon. Junction sequences derived from patients with leukemias carrying t(4:11), t(11:17), or t(9:11) abnormalities were PCR-amplified and used to create pSG5 ALL-1/AF-4, ALL-1/AF-17, or ALL-1/AF9/FLAG constructs.

An expression vector for chicken muscle pyruvate kinase (PK), RLPK12, was described elsewhere (28, 29). We incorporated a FLAG peptide-encoding sequence at the C-terminal end of the PK cDNA sequence (RLPK12-FLAG) so that the ALL-1 polypeptides linked in-frame to PK could be identified with mouse anti-FLAG mAb. Segments derived from ALL-1 cDNA were PCR-amplified or purified electrophoretically after digestion of bigger fragments, and subcloned at *Xho*I and *Eco*RI sites incorporated at the N terminus of the PK coding sequence.

To create constructs with mutations or deletions, site-directed mutagenesis was performed using Ex-site PCR-based site-directed mutagenesis kit (Stratagene) with slight modifications.

Transfection and Immunostaining. COS, HeLa, or SV80 cells were grown in six-well culture dishes on 18-mm² glass coverslips in DMEM supplemented with 10% fetal bovine serum for 16–30 hr to 80% confluency. For transfection, cells were incubated with lipofectamine (GIBCO/BRL) containing 2 μ g of plasmid DNA per well, according to the manufacturer's descriptions. Thirty-six hours after transfection, cells were washed in PBS and fixed in methanol at -20°C for 10 min. Fixed cells were incubated with normal blocking serum in a humid chamber at room temperature and subsequently with mouse anti-FLAG M2 mAb (Eastman Kodak), or with rabbit anti-human ALL-1 antibody (Ab) (the P4 and P3 Ab were raised against peptides spanning residues 1,281–1,299 and 153–167, respectively) at concentrations of 15 μ g/ml for 1 hr. The cells were washed with PBS for 15 min and finally incubated for 1 hr in a mixture containing TRITC (tetramethyl-rhodamin isothiocyanate)-conjugated goat anti-mouse IgG (Jackson ImmunoResearch) diluted 1:300 and the DNA-binding fluorescent dye DAPI (4',6-diamidino-2-phenylindole; 1 μ g/ml). After a 15-min wash with PBS, cells were embedded in elvanol and images of labeled cells were recorded using Zeiss Axiophot fluorescent microscope with a Plan \times 100 objective.

Electron Microscopy Analysis. Preparation of the samples and immunolabeling were performed using modifications of the Tokuyasu method as previously described (30). The cells were fixed with 0.1% glutaraldehyde and 3% formaldehyde in 0.1 M cacodylate buffer (pH 7.2) for 1 hr. The cells were washed in the buffer, scraped off the plate, and embedded in 10% gelatin. The gelatin blocks were postfixated as above for 16 hr, incubated in 2.3 M sucrose, and cryosectioned at -120°C using a Reichert FCS ultracytome. The ultrathin sections

were collected on grids, washed in PBS, and blocked with 0.5% BSA, 1% normal goat serum, 1% gelatin, 1% cystein, and 1% Tween-20 in PBS. The sections were immunolabeled as indicated above, embedded in methylcellulose, and examined with a Philips model EM410 electron microscope at 80 kV.

RESULTS

Endogenous ALL-1 and Chimeric ALL-1 Proteins Show a Speckled Nuclear Distribution. Utilizing two antipeptide Abs, we detected the normal ALL-1 protein in the nucleus with no significant staining in the cytoplasm (Fig. 1). The protein was excluded from the nucleolus and showed a pattern of small speckles uniformly distributed in the nucleoplasm. The same pattern was observed in COS, HeLa, and SV80 cells, as well as in a series of leukemic cell lines with or without 11q23 abnormalities. We now investigated the cellular distribution of

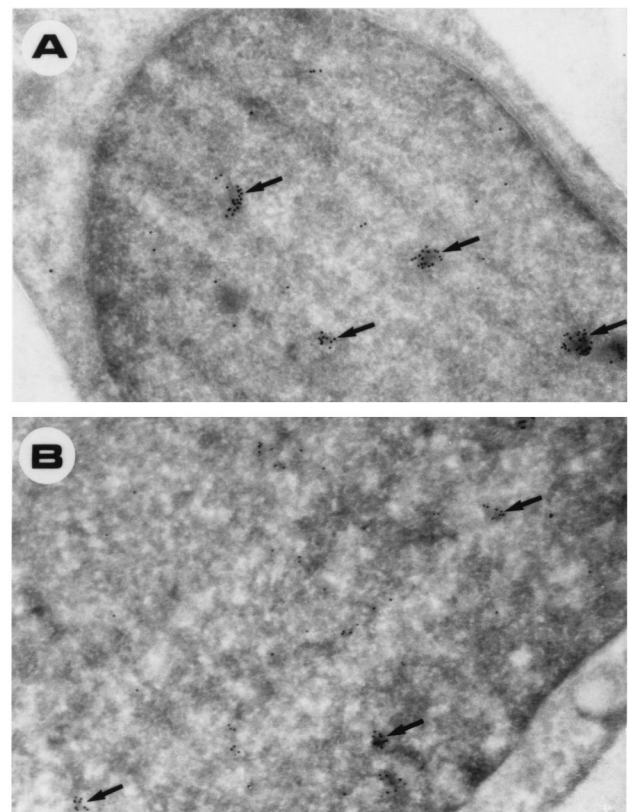


FIG. 2. (A) Immunogold labeling of COS cells for endogenous ALL-1 by using rabbit P4 Ab followed by gold (15 nm)-conjugated goat anti-rabbit IgG (Zymed). (\times 42,500.) (B) Immunogold labeling of COS cells transfected with FLAG-tagged ALL-1/AF-9 utilizing monoclonal mouse antibody to the epitope and gold (10 nm)-conjugated goat anti-mouse IgG (Zymed). (\times 42,500.)

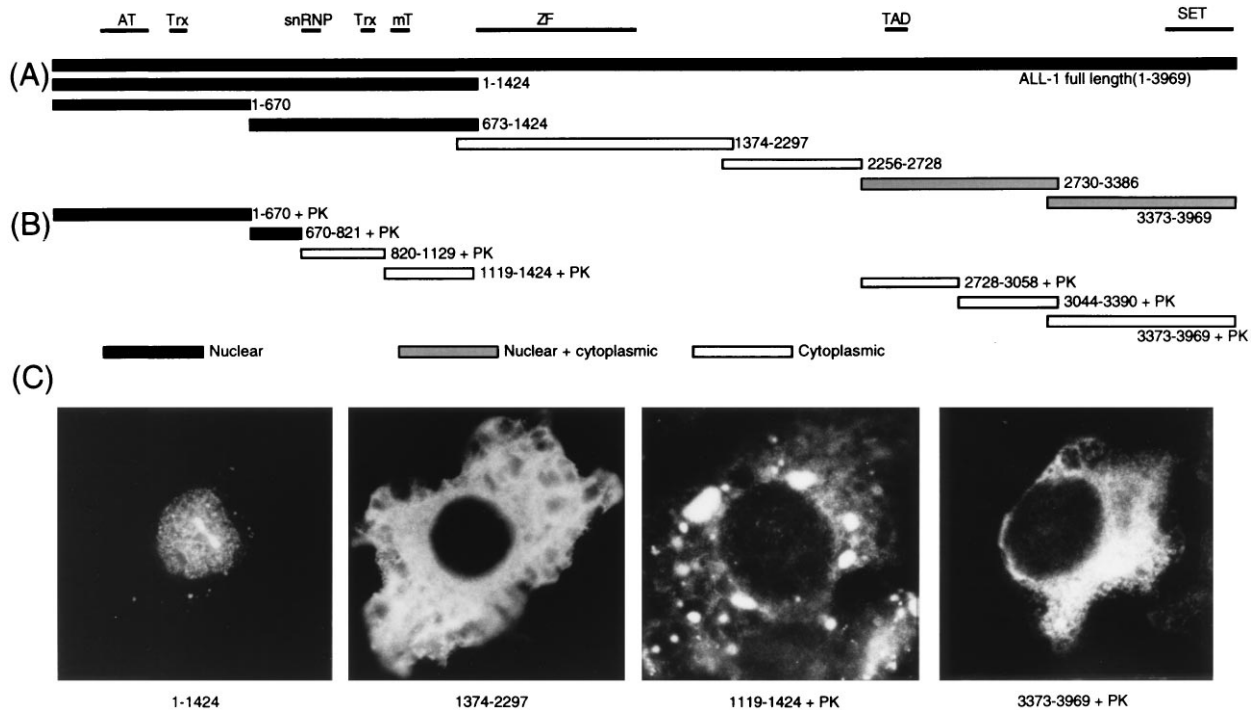


FIG. 3. Mapping nuclear targeting signals of ALL-1. Protein segments were linked to the FLAG tag, inserted into the pSG5 vector (A) or the RLPK12 vector (B) and transiently transfected into COS cells. (C) Representative stained cells. AT and ZF correspond to AT hook and zinc finger (PHD finger) domains, respectively.

chimeric ALL-1 proteins produced in acute leukemias with the t(4;11), t(9;11), or t(11;17) abnormalities. The proteins were transiently overexpressed in COS cells and detected by staining with anti-ALL-1 polyclonal Ab. Being tagged with the FLAG peptide, ALL-1/AF-9 was also detected with anti-FLAG mAb. The three proteins showed nuclear punctate pattern (Fig. 1). The pattern was less uniform than that of endogenous ALL-1, and consisted of small dots and bigger patches. The proteins were present within both the nucleoli and nucleoplasm and sometimes at the nucleus' periphery. The difference in distribution of the endogenous and exogenous ALL-1 proteins might be due to overexpression of the latter. Although

overexpressed truncated ALL-1 (amino acids 1–1,424) showed a pattern comparable to that of the chimeric proteins (see below), we cannot exclude some effects of the partner polypeptides on the distribution.

In an attempt to further localize the ALL-1 proteins, we used electron microscopy (Fig. 2). The staining was predominantly nuclear, mostly organized in the form of small clusters with a typical diameter of about 1 micron. The exogenous molecules often distributed as less homogenous clusters. The labeling was rarely associated with nucleoli and was not preferentially enriched near the lamina.

Identification of ALL-1 Elements Directing Nuclear Localization. Segments of the ALL-1 protein were epitope-tagged with the FLAG peptide. The constructs were inserted into the expression vector pSG5 and transfected into COS cells, and the FLAG-tagged proteins examined for subcellular distribution (Fig. 3A). Fragments from the N terminus and the center of the protein were found present in the nucleus and cytoplasm, respectively. Fragments of the C terminus were found associated with the cytoplasm or with the entire cell. For further confirmation and/or identification of ALL-1 fragments conferring nuclear localization, we used the established methodology (28, 29) of linking in-frame small polypeptides of the studied protein to the cytoplasmic enzyme PK and determining the subcellular distribution of the fused proteins (Fig. 3B). Examples of this and the previous analysis are shown in Fig. 3C. The results indicate that ALL-1 N-terminal ≈ 820 residues contain nuclear targeting sequences. Interestingly, the region with homology to mT confers localization in cytoplasmic speckles positioned at the periphery of the nucleus; no other ALL-1 segment directed such a pattern.

The N-terminal region of ALL-1 was further dissected (Fig. 4) to identify three relatively small polypeptides, NTS-1, NTS-2, and NTS-3 (NTS, nuclear targeting sequence) directing PK to the nucleus. NTS-1 was less efficient than the other two, so that 10–15% of the staining was detected in the cytoplasm. NTS-3 spans a classical nuclear targeting sequence, RKRKRK, positioned at residues 734–739. We note that an NTS-3 mutant with a deletion of the six basic amino acids

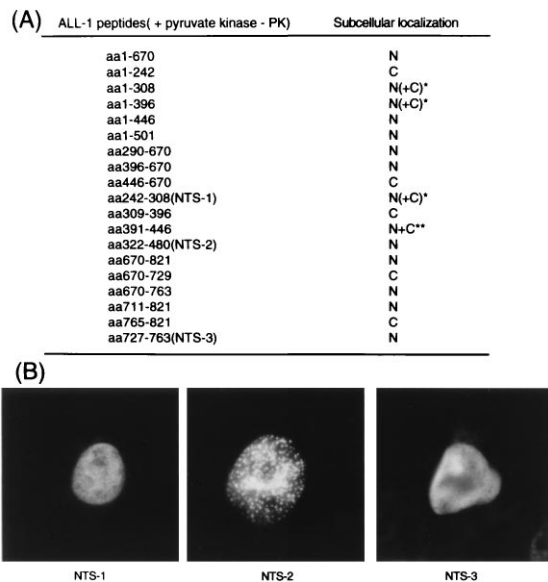


FIG. 4. (A) Fine mapping of ALL-1 nuclear targeting signals. *, Ninety percent of staining in nucleus; **, about equal nuclear and cytoplasm staining. (B) Typical patterns of cells expressing NTS-1, -2, or -3.

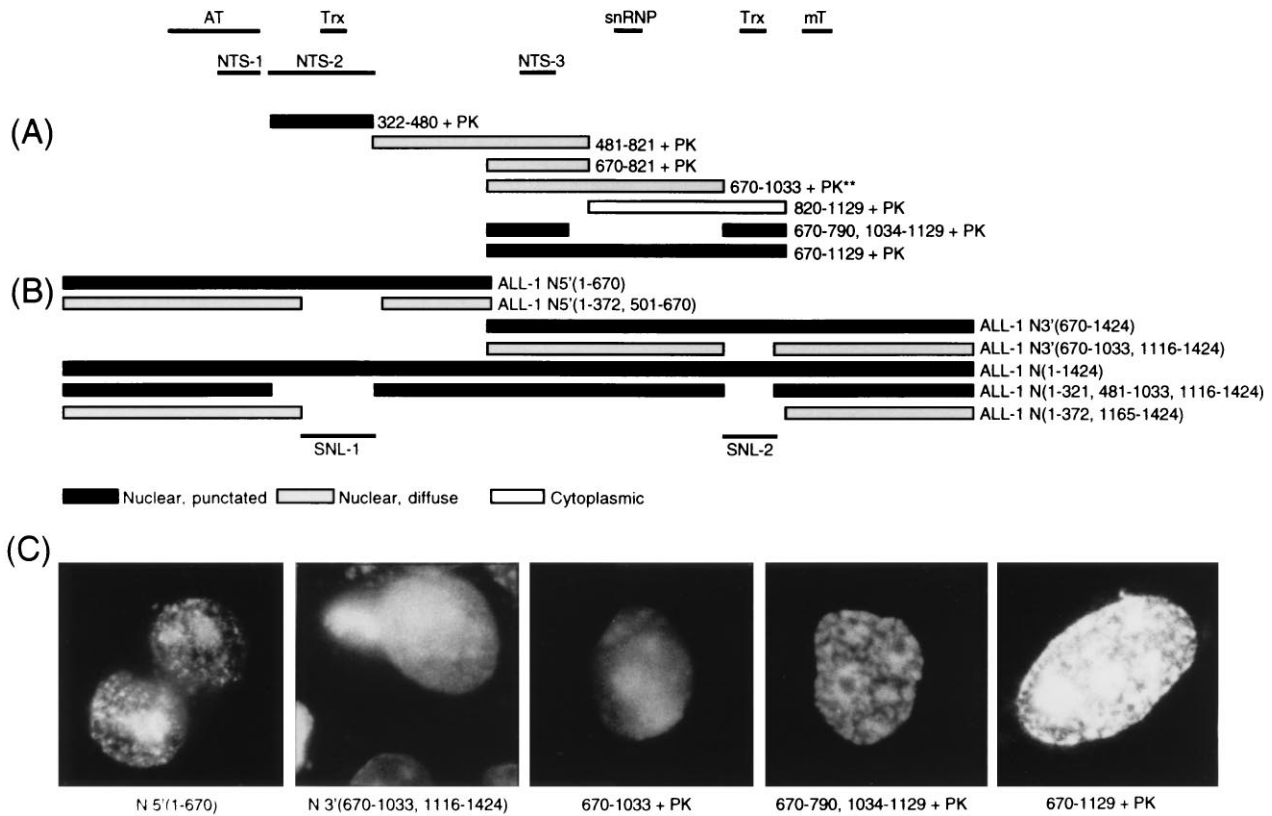


FIG. 5. ALL-1 segments conferring punctate nuclear pattern, inserted within the RLPK12 vector (A) or the pSG5 vector (B). **, A minority of cells show unusual, poorly defined patches. (C) Representative examples.

(734–739) was incapable of nuclear targeting when tethered to PK (data not shown).

Elements Mediating Punctate Pattern. Polypeptides containing NTS-2 localized as nuclear speckles (Fig. 4B), reminiscent of endogenous and exogenous ALL-1 proteins. By testing a series of constructs containing PK linked to small segments of ALL-1, we determined two domains conferring punctate distribution (Fig. 5A). SNL-1 (SNL, speckled nuclear localization) co-maps with NTS-2. In contrast, SNL-2 did not determine nuclear localization by itself, but once there (through linking to NTS-3 or NTS-1) it would appear in speckles. Remarkably, SNL-1 and SNL-2 spanned short ALL-1 domains (*Trx* in Fig. 5)—amino acids 388–432 and 1,051–1,089, respectively—which are conserved with TRX (24, 25) and are nearly identical in the latter and in the homologous protein from *Drosophila virilis* (25). Deletions of SNL-1 and SNL-2 from the 5' and 3' halves of truncated ALL-1, respectively, indeed resulted in nuclear nonpunctate distribution (Fig. 5B). While SNL-1 and SNL-2 were the only domains identified by the PK methodology, as domains conferring dots, we found that cells expressing an ALL-1 N (amino acids 1–1,424) protein from which the two domains were deleted (residues 322–480 and 1,034–1,115) still exhibited nuclear speckles (Fig. 5B). Thus, other N-terminal determinants could still confer a dot pattern. These determinants appear to be included within residues 373–1,164 because deletion of the segment abolished punctate distribution completely (Fig. 5B).

Deletion of amino acids 408–432 from the NTS-2/SNL-1 (amino acids 322–480)-PK construct resulted in a diffuse and predominantly (80–90% of the staining) nuclear distribution of the protein (Fig. 6A). An NTS-2/SNL-1-PK construct, mutated at amino acids 418–423 (SSRIIK to AAAAAA), which are fully conserved between ALL-1 and TRX, showed an ≈ 60 –70% decrease in the number of cells displaying a punctate pattern, although the protein remained exclusively in the nucleus (Fig. 6A). Moreover, PK constructs spanning

ALL-1 residues 322–407 or 433–480 distributed in the cytoplasm and predominantly in the nucleus, respectively (data not shown). This indicates that within the NTS-2/SNL-1 region, speckle formation is solely conferred by residues 408–432 conserved with TRX, while the major nuclear targeting activity is directed by amino acids 433–480. Fine deletion/mutagenesis methodology was also applied to the SNL-2 domain (Fig. 6A). Deletion of the central TRX conserved region (amino acids 1,065–1,089) from the SNL-2 (amino acids 1,034–1,115)-PK protein resulted in loss of the punctate pattern. Alterations of the conserved residues 1,065–1,073 (GPRIKHVCR to AAAAAAAA) reduced by ≈ 80 % the number of cells with dots.

The dichotomy between two ALL-1 elements directing nuclear localization (NTS-1 and NTS-3) and an element conferring punctate pattern (SNL-2) enabled us to investigate whether other domains of the ALL-1 protein, derived from the center and C terminus of the molecule, when artificially expressed in the nucleus would distribute in speckles. Thus, ALL-1 segments spanning either the polypeptide with homology to mT, or the PHD fingers region, or the transcriptional TAD, or the SET domain were fused to NTS-1 or NTS-3 and examined for subnuclear distribution (Fig. 6B). None of these segments directed a punctate pattern. This indicates that only ALL-1 N-terminal elements are capable of interactions resulting in subnuclear foci distribution.

DISCUSSION

The two main findings reported here are the presence of the ALL-1 protein in multiple large subnuclear structures and the identification of several elements within the N-terminal one-third of the molecule, responsible for nuclear localization and/or distribution into speckles. Overexpressed leukemia-associated chimeric ALL-1 proteins or overexpressed specific ALL-1 segments were also present as nuclear dots. These dots

1. Gu, Y., Nakamura, T., Alder, H., Prasad, R., Canaani, O., Cimino, G., Croce, C. M. & Canaani, E. (1992) *Cell* **71**, 701–708.
2. Tkachuk, D., Kohler, S. & Cleary, M. (1992) *Cell* **71**, 691–700.
3. Raimondi, S. C., (1993) *Blood* **81**, 2237–2251.
4. Schichman, S. A., Canaani, E. & Croce, C. M. (1995) *J. Am. Med. Assoc.* **273**, 571–576.
5. Canaani, E., Nowell, P. C. & Croce, C. M. (1995) *Adv. Cancer Res.* **66**, 213–234.
6. Bernard, O. A. & Berger, R. (1995) *Genes Chromosomes Cancer* **13**, 75–85.
7. Corral, J., Lavenir, I., Impey, H., Warren, A. J., Forster, A., Larson, T. A., Bell, S., McKenzie, A. N. J., King, G. & Rabbits, T. H. (1996) *Cell* **85**, 853–861.
8. Djabali, M., Selleri, L., Parry, P., Bower, M., Young, B. D. & Evans, G. A. (1992) *Nat. Genet.* **2**, 113–118.
9. Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A. J. & Korsmeyer, S. J. (1995) *Nature (London)* **378**, 505–508.
10. Lewis, E. B. (1978) *Nature (London)* **276**, 565–570.
11. Kennison, J. A. (1995) *Annu. Rev. Genet.* **29**, 289–303.
12. Orlando, V. & Paro, O. (1995) *Curr. Opin. Genet. Dev.* **5**, 174–179.
13. Franke, A., De Camillis, M., Zink, D., Cheng, N., Brock, H. W. & Paro, R. (1992) *EMBO J.* **11**, 2941–2950.
14. Lonie, A., D'Andrea, R., Paro, R. & Saint, R. (1994) *Development (Cambridge, U.K.)* **120**, 2629–2636.
15. Rastelli, L., Chan, C. S. & Pirota, V. (1993) *EMBO J.* **12**, 1513–1522.
16. Kuzin, B., Tillib, S., Sedkov, Y., Mizrokhi, L. & Mazo, A. (1994) *Genes Dev.* **8**, 2478–2490.
17. Tripoulas, N., La Jeunesse D., Gildea, J. & Shearn, A. (1996) *Genetics* **143**, 913–928.
18. Chinwalla, V., Jane, E. P. & Harte, P. J. (1995) *EMBO J.* **14**, 2056–2065.
19. Kwon, H., Imbalanzo, A. N., Khavari, P. A., Kingston, R. E. & Green, M. R. (1994) *Nature (London)* **370**, 477–481.
20. Peterson, C. L. & Tamkun, J. W. (1995) *Trends Biochem. Sci.* **20**, 143–146.
21. Tsukiyama, T., Becker, P. B. & Wu, C. (1994) *Nature (London)* **367**, 525–532.
22. Tschiersch, B., Hofmann, A., Krauss, V., Dorn, R., Korge, C. & Reuter, G. (1994) *EMBO J.* **13**, 3822–3831.
23. Jones, R. S. & Gelbart, W. M. (1993) *Mol. Cell. Biol.* **13**, 6357–6366.
24. Domer, P. H., Frakharzadeh, S. S., Chen, C.-H., Jackel, J., Johansen, L., Silverman, G. A., Kersey, J. H. & Korsmeyer, S. J. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7884–7888.
25. Tillib, S., Sedkov, Y., Mizrokhi, L. & Mazo, A. (1995) *Med. Dev.* **53**, 113–122.
26. Stassen, M. J., Bailey, D., Nelson, S., Chinwalla, V. & Harte, P. J. (1995) *Med. Dev.* **52**, 209–223.
27. Prasad, R., Yano, T., Sorio, C., Nakamura, T., Rallapalli, R., Gu, Y., Leshkowitz, D., Croce, C. M. & Canaani, E. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 12160–12164.
28. Kalderon, D., Roberts, B. L., Richardson, W. D. & Smith, A. (1984) *Cell* **39**, 499–509.
29. Dang, C.-V. & Lee, W. M. F. (1988) *Mol. Cell. Biol.* **8**, 4048–4054.
30. Geiger, B., Dutton, A. H., Tokuyasu, K. T. & Singer, S. J. (1981) *J. Cell Biol.* **91**, 614–628.
31. Alkema, M. J., Bronk, M., Verhoeven, E., Otte, A., van Veer, L. J., Berns, A. & Van Lohuizen, M. (1997) *Genes Dev.* **11**, 226–240.
32. Fu, X.-D. & Maniatis, T. (1990) *Nature (London)* **343**, 437–441.
33. O'Keefe, R. T., Mayeda, A., Sadowski, C. L., Krainer, A. R. & Spector, D. L. (1994) *J. Cell Biol.* **124**, 249–260.
34. Moir, R. D., Montag-Lowg, M. & Goldman, R. D. (1994) *J. Cell Biol.* **125**, 1201–1212.
35. Kill, I. R., Bridger, J. M., Campbell, K. H. S., Maldonado-Codina, G. & Hutchinson, C. J. (1991) *J. Cell Sci.* **100**, 869–876.
36. Weis, K. I., Rambaud, S., Lavau, C., Jansen, T., Carvalho, T., Carmo, F. M., Lamond, A. & Dejean, A. (1994) *Cell* **76**, 345–356.
37. Dyck, J. A., Maul, G. C., Miller, W. H., Chen, J. D., Kakikuzo, A. & Evans, R. M. (1994) *Cell* **76**, 333–343.
38. Scully, R., Chen, J., Plug, A., Xiao, Y., Weaver, D., Feunteun, J., Ashley, T. & Livingston, D. M. (1997) *Cell* **88**, 265–275.
39. Messner, S., Franke, A. & Paro, R. (1992) *Genes Dev.* **6**, 1241–1254.