

on fluorescence in situ hybridization (FISH) ordering of regionally derived cosmid clones (5). Bidirectional chromosome walks were performed from cosmids ~290 kilobases (kb) apart that flanked the breakpoint on chromosome 5; each walk spanned a genomic region of 150 kb. With genomic probes isolated 70 kb from the telomeric cosmid (6), we observed rearranged restriction fragments in the DNAs of two lymphoma cell lines containing the t(2;5) (Fig. 1A). One of the probes (p21-3/3E) was hybridized to a complementary DNA (cDNA) library prepared from the UOC-B1 pro-B leukemia cell line, which is t(2;5)-negative. Multiple cDNA clones were isolated that hybridized to a ubiquitously expressed 1.6-kb mRNA encoding nucleophosmin (NPM; also known as B23 or numatrin), a highly conserved nucleolar phosphoprotein that shuttles ribosomal components between the nucleolus and the cytoplasm in the later stages of ribosome assembly (7). Northern (RNA) blot analysis with a probe corresponding to 5'-specific NPM sequences revealed that t(2;5)-positive cell lines expressed both the normal 1.6-kb NPM transcript and a 2.4-kb transcript, whereas t(2;5)-negative cell lines expressed only the 1.6-kb transcript (Fig. 1B). In contrast, a probe containing 3'-untranslated NPM sequences detected only the normal 1.6-kb NPM transcript in all cell lines (8).

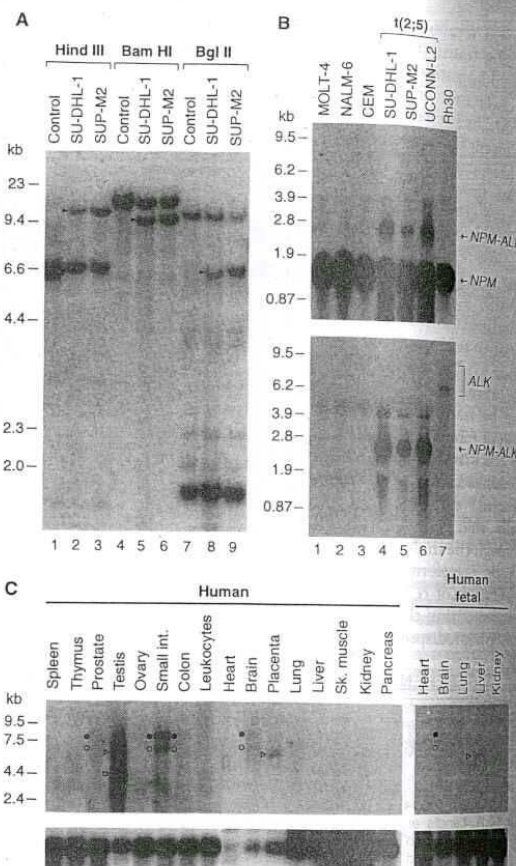
By screening a cDNA library prepared from the SU-DHL-1 t(2;5)-positive cell line, we isolated more than 20 clones that hybridized to 5' but not 3' NPM probes. Sequences from the 5' ends of the three longest clones were identical to 5' NPM cDNA sequences but diverged after the codon for Val¹¹⁷. NPM sequences 3' of this codon were replaced by 1223 nucleotides (nt), resulting in an open reading frame of 1575 nt (Fig. 2A). A probe (pS1.2) representing the 3' end of the fusion cDNA identified the same 2.4-kb transcript that had been detected with the 5'-specific NPM probe in RNAs from t(2;5)-positive cells (Fig. 1B). This fragment was localized to chromosome 2p23 by hybridization to DNAs of human-rodent somatic cell hybrids and by metaphase FISH analysis (9), which indicates that the 2.4-kb mRNA was encoded by a fused gene created by the t(2;5).

The 3' portion of the chimeric t(2;5) cDNA encodes conserved residues characteristic of the catalytic domain of protein tyrosine kinases (PTKs) (10) (Fig. 2, A and B). The newly identified anaplastic lymphoma kinase (ALK) showed the greatest sequence similarity to members of the insulin receptor kinase subfamily, including leukocyte tyrosine kinase (LTK; 64% amino acid identity), TRKA (38%), ROS (37%)

and its *Drosophila* homolog Sevenless (35%), the β chain of the insulin-like growth factor receptor (IGF-1 receptor; 37%), and the β chain of the insulin receptor (IR; 36%). We determined the structure of normal ALK proteins by isolating cDNA clones from the Rh30 rhabdomyosarcoma cell line with the pS1.2 probe. Analysis of the inserts of the two largest clones, pRMS4 and pRMS17-2, revealed 3' ALK sequences identical to those in the fusion gene cDNA, which indicates that mutations had not occurred in the chimeric protein. Sequences of ALK immediately upstream of the NPM-ALK junction encoded 23 hydrophobic amino acids typical of a transmembrane domain, whereas those from the extreme 5' ends of the ALK clones were 50% identical to sequences encoding insulin-like growth factor binding protein (IBP-1) (8, 11). These comparisons indicate that the normal ALK product is a membrane-span-

ning tyrosine kinase receptor. Notably, the transmembrane and putative extracellular domains are not included in the NPM-ALK chimeric protein.

ALK mRNAs of 6.5 and 8.0 kb were abundantly expressed in rhabdomyosarcoma cell lines and in small intestine and were weakly expressed in brain (fetal and adult), colon, and prostate (Fig. 1, B and C). Large amounts of 4.4-kb and 6.0-kb mRNAs were detected in testis, whereas only a 6.0-kb mRNA was detected in placenta and fetal liver. All four mRNAs were also detected with a probe containing only 3' untranslated ALK sequences, which suggests that they represent differentially spliced ALK mRNAs, not cross-hybridizing transcripts of other PTK genes. ALK transcripts were not detected in Northern analysis of RNA prepared from hematopoietic cells, including normal spleen, thymus, peripheral blood leukocytes, B lymphoblastoid cell lines, phytohemagglutinin-stim-



and its *Drosophila* homolog Sevenless (35%), the β chain of the insulin-like growth factor receptor (IGF-1 receptor; 37%), and the β chain of the insulin receptor (IR; 36%).

We determined the structure of normal ALK proteins by isolating cDNA clones from the Rh30 rhabdomyosarcoma cell line with the pS1.2 probe. Analysis of the inserts of the two largest clones, pRMS4 and pRMS17-2, revealed 3' ALK sequences identical to those in the fusion gene cDNA, which indicates that mutations had not occurred in the chimeric protein. Sequences of ALK immediately upstream of the NPM-ALK junction encoded 23 hydrophobic amino acids typical of a transmembrane domain, whereas those from the extreme 5' ends of the ALK clones were 50% identical to sequences encoding insulin-like growth factor binding protein (IBP-1) (8, 11). These comparisons indicate that the normal ALK product is a membrane-span-

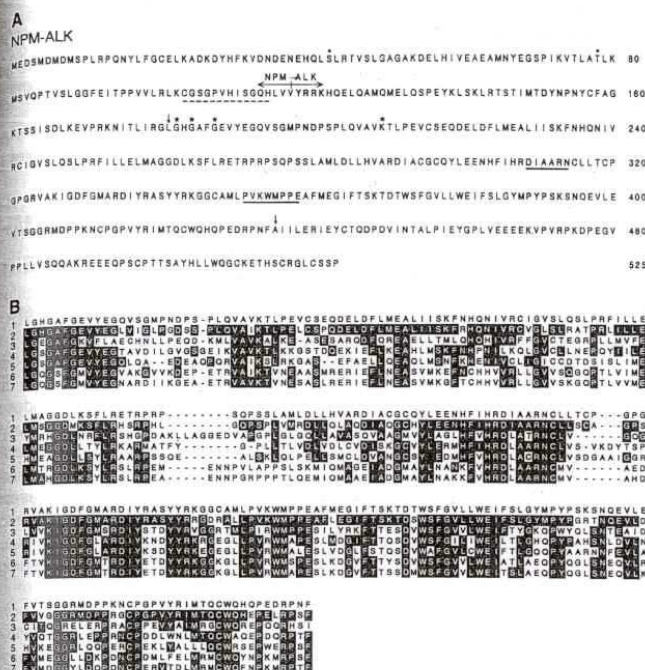


Fig. 2. (A) Deduced amino acid sequence of NPM-ALK (20). Solid circles, potential phosphorylation sites for protein kinase C; dashed underline, potential metal binding domain; arrows, boundaries of the ALK catalytic domain; asterisks, conserved residues of the consensus adenosine triphosphate (ATP) recognition sequence and the ATP-binding lysine residue; and solid underlines, consensus sequences specific for tyrosine kinases. (B) Alignment of the ALK catalytic domain with those in the insulin receptor subfamily of PTKs. Line 1, ALK; line 2, LTK; line 3, TRKA; line 4, ROS; line 5, Sevenless; line 6, IGF-1 receptor; and line 7, IR. Gaps are indicated by dashes, and amino acid identity is indicated by shaded boxes. All sequences are for human proteins, except for Sevenless, which is a *Drosophila* protein (10).

lated T lymphocytes, or t(2;5)-negative leukemia-lymphoma cell lines of myeloid or B or T lymphoid derivation.

FISH mapping indicated that NPM and ALK are transcribed in centromeric to telomeric orientations on chromosomes 5 and 2, respectively, with the 2.4-kb fusion transcript arising from the derivative 5 translocated chromosome. Northern blot analysis provided no evidence for expression of a reciprocal ALK-NPM chimeric transcript, which could have been generated from the derivative 2 chromosome.

An RNA-based polymerase chain reaction (RNA-PCR) method (12) confirmed the specificity of the fusion junctions in chimeric transcripts expressed in lymphomas having the t(2;5) (Fig. 3). Conversely, fusion transcripts were not detected in t(2;5)-negative cell lines, including several rhabdomyosarcoma lines that expressed ALK transcripts. Identical NPM-ALK junction sequences were found in the RNAs of all seven t(2;5)-positive samples, including the SU-DHL-1, SUP-M2, and UCONN-

L2 cell lines and diagnostic samples from four patients with anaplastic large-cell lymphomas (13). The breakpoints of the 2;5 translocation therefore appear to consistently involve the same introns of the NPM and ALK genes, leading to identical junctions in the fusion mRNAs.

The normal NPM protein is a nonribosomal nucleolar phosphoprotein involved in the assembly of preribosomal particles into both small and large ribosomal subunits (7). It binds cooperatively with high affinity to single-stranded nucleic acids, exhibits activity that destabilizes the RNA helix, and is found in association with the most mature nucleolar preribosomal ribonucleoproteins (14). NPM transcription and translation are cell cycle-regulated, reaching peak levels just before the entry of the cells into the S phase, with a decline to the base line just before the onset of the G₂ phase (15). We postulate that the NPM gene contributes an active promoter to drive expression of the ALK catalytic domain in lymphoma cells containing the

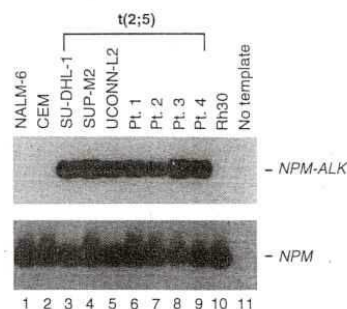


Fig. 3. Southern blot analysis of NPM-ALK and NPM RNA-PCR products. Total RNAs (1 μ g) from t(2;5)-positive cell lines (SU-DHL-1, SUP-M2, and UCONN-L2) and patient (Pt.) samples (1 to 4) were analyzed. RNAs from the t(2;5)-negative B and T lymphoid leukemia cell lines (NALM-6 and CEM, respectively) and the Rh30 rhabdomyosarcoma cell line, which expresses normal ALK, were included as negative controls, as was a blank without an RNA template.

t(2;5). This role for NPM would appear to be crucial, because the ALK promoter is normally silent in lymphoid cells. An oncogenic role, if any, remains to be established for the NH₂-terminal NPM coding sequences incorporated into NPM-ALK, including those encoding potential protein kinase C phosphorylation sites (Ser⁴³ and Thr⁷⁸) and a potential Cys-X₅-His-X₄-His metal binding motif (residues 104 to 115, where X is any amino acid) (7).

A consistent feature of tyrosine kinase oncogenes, including TRKA, BCR-ABL, EGFR, ERBB2, CSF-1R, and LCK, is that much of their potency can be attributed to mutations or gene fusions that lead to a constitutively active catalytic domain (16). Thus, in NPM-ALK fusion proteins, one would predict that the truncated ALK is deregulated and phosphorylates intracellular substrates to trigger malignant transformation. Anaplastic large-cell lymphomas arise from activated T lymphocytes, which depend on IL-2 for growth and viability (17). Conceivably, NPM-ALK phosphorylates substrates that are normally phosphorylated in response to IL-2 receptor- or T cell receptor-mediated signals (18), leading to constitutive activation of one of these signal transduction pathways.

REFERENCES AND NOTES

1. H. Stein and F. Dallenbach, in *Neoplastic Hematopathology*, D. M. Knowles, Ed. (Williams & Wilkins, Baltimore, 1992), pp. 675-714.
2. D. Y. Mason et al., *Br. J. Haematol.* 74, 161 (1990); M. A. Bitter, *Am. J. Surg. Pathol.* 14, 305 (1990); M. E. Kadin, *J. Clin. Oncol.* 9, 533 (1991); J. P. Greer et al., *ibid.*, p. 539.
3. H. Durkop et al., *Cell* 68, 421 (1992); C. A. Smith et al., *ibid.* 73, 1349 (1993).
4. A. G. Stansfeld et al., *Lancet* 1, 292 (1988).
5. To identify the breakpoint on chromosome 5, we

isolated microdissection clones from bands 5q34-q35 and used them to identify 39 cosmid clones [D. Saltman *et al.*, *Nucleic Acids Res.* 20, 1401 (1992)], which then were oriented by FISH analysis of metaphase chromosomes from the SUP-M2 and SU-DHL-1 t(2;5)-positive cell lines [R. Morgan *et al.*, *Blood* 73, 2155 (1989)]. Seventeen clones mapped centromeric and 22 clones telomeric to the breakpoint; clones from these groups were oriented relative to one another by two-color metaphase FISH analysis. The estimated genomic distance between the two cosmids that flanked the breakpoint most closely was 290 kb by interphase FISH analysis [J. B. Lawrence, R. H. Singer, J. A. McNeil, *Science* 249, 928 (1990); B. Trask *et al.*, *Am. J. Hum. Genet.* 48, 1 (1991)]. Probes prepared from these cosmids did not detect rearranged restriction fragments by Southern (DNA) blot analysis of pulsed-field gels containing DNA from t(2;5)-positive cell lines.

6. The genomic fragment p16-3/1.2S is located immediately centromeric to the chromosome 5 breakpoint, whereas fragment p21-3/3E lies just telomeric to the break. Both probes identified a 1.6-kb transcript in Northern analysis of RNAs prepared from t(2;5)-positive and t(2;5)-negative cell lines; in addition, p16-3/1.2S hybridized to a 2.4-kb transcript found only in t(2;5)-positive cells.
7. M. S. Schmidt-Zachmann and W. W. Franke, *Chromosoma* 96, 417 (1988); W. Y. Chan *et al.*, *Biochemistry* 28, 1033 (1989); R. A. Borer, C. F. Lehner, H. M. Eppenberger, E. A. Nigg, *Cell* 56, 379 (1989).
8. S. W. Morris, unpublished data.
9. _____ and M. B. Valentine, unpublished data.
10. S. K. Hanks, A. M. Quinn, T. Hunter, *Science* 241, 42 (1988).
11. IBP-1 is a 30-kD secreted protein found in human plasma and amniotic fluid that binds IGF-1 with high affinity [A. Brinkman, C. Groffen, D. J. Kortleve, A. Geurts van Kessel, S. L. Drop, *EMBO J.* 7, 2417 (1988)].
12. RNA-PCR reactions were performed simultaneously with oligonucleotide primers specific for the chimeric *NPM-ALK* transcript (5' *NPM*: 5'-TCCCTTG-GGGGCTTTGAAATACACC-3'; and 3' *ALK*: 5'-CG-AGGTGCGGAGCTTGGCTCAGC-3') and with a primer pair derived from the ubiquitously expressed *NPM* gene as a control for reverse transcription and amplification (5' *NPM* and 3' *NPM*: 5'-GCTACCACTCCAGGGGCGAGA-3'). The 177-bp *NPM-ALK* product was detected by hybridization with an end-labeled oligonucleotide homologous to sequences spanning the fusion junction (5'-AGCACTTAGTAG-TGTACCGCCGGA-3'); the 185-bp *NPM* product was detected with an oligonucleotide homologous to normal *NPM* sequences in the region of the junction (5'-AGCACTTAGTAGCTGTGGAGG-AAG-3').
13. Written informed consent was obtained from the patients or their parents, and investigations were approved by the clinical trials review committee of St. Jude Children's Research Hospital.
14. T. S. Dumbear, G. A. Gentry, M. O. J. Olson, *Biochemistry* 28, 9495 (1989).
15. N. Feuerstein, S. Spiegel, J. J. Mond, *J. Cell Biol.* 107, 1629 (1988); N. Feuerstein and P. A. Randazzo, *Exp. Cell Res.* 194, 289 (1991).
16. J. Schlessinger and A. Ullrich, *Neuron* 9, 383 (1992); T. Pawson, *Curr. Opin. Gen. Dev.* 2, 4 (1992); R. C. Burnett *et al.*, *Genes Chrom. Cancer* 3, 461 (1991); B. Tycko, S. D. Smith, J. Sklar, *J. Exp. Med.* 174, 867 (1991).
17. K. A. Smith, *Science* 240, 1169 (1988).
18. I. D. Horak *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 88, 1996 (1991); R. D. Klausner and L. E. Samelson, *Cell* 64, 875 (1991); N. Kobayashi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 90, 4201 (1993).
19. The faint ~4-kb RNAs detected by pS1.2 in the t(2;5)-positive cells represent cross-hybridization of the probe with the 28S ribosomal RNA.
20. The nucleotide sequence of *NPM-ALK* (accession number U04946) has been deposited in GenBank. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro;

Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

21. We thank A. Sinclair and J. Sublett for technical assistance; M. Lovett for support and advice; L. Deaven for the chromosome 5 cosmid library; J. Gilbert for editorial review; PC&T for their unwavering support; K. Williams, J. T. Sandlund, F. Behm, D. Head, and J. Whitlock for help with clinical lymphoma samples; and J. Cleveland, J.

Downing, V. Kidd, J. Ihle, A. Nienhuis, and C. Sherr for critical review of the manuscript. Supported by National Cancer Institute grants K08 CA 01702, P01 CA 20180, and Core grant CA 21765, and by the American Lebanese Syrian Associated Charities of St. Jude Children's Research Hospital.

23 September 1993; accepted 18 January 1994

Diversity of Endogenous Epitopes Bound to MHC Class II Molecules Limited by Invariant Chain

Helen Bodmer,* Stéphane Viville,† Christophe Benoist, Diane Mathis

The invariant chain (Ii) binds nascent major histocompatibility complex (MHC) class II molecules, blocking peptide binding until the complex dissociates in the endosomes. This may serve to differentiate the MHC class I and II antigen presentation pathways and enable class II molecules to efficiently bind peptides in the endosomes. This hypothesis was addressed by probing spleen cells from a combination of knock-out and transgenic mice with a large panel of T cell hybridomas. The Ii molecule blocked the presentation of a range of endogenously synthesized epitopes, but some epitopes actually required Ii. Thus, the influence of Ii on presentation does not follow simple rules. In addition, mice expressing Ii were not tolerant to epitopes unmasked in its absence, a finding with possible implications for autoimmunity.

The primary function of MHC class II molecules is to present peptides to CD4⁺ T cells, often those derived from foreign, endocytosed proteins and encountered by class II molecules as they traffic through endocytic vesicles (1). In theory, peptides from exogenous antigen must compete for presentation with the vast quantities of endogenous proteins that transit through the endoplasmic reticulum (ER) and Golgi compartments and are known to engender MHC-binding peptides (1–3). It has been hypothesized that efficient loading of class II molecules in the ER and Golgi is prevented by Ii, a transmembrane protein that targets class II molecules to endosomes (4–6) and interferes with MHC class II-peptide interactions (7). However, this hypothesis has been challenged as unnecessary because many class II-peptide interactions are inefficient at the neutral pH of the ER and Golgi, requiring the more acidic conditions of the endosomes (8). In addition, recent studies on mice lacking Ii showed that, in the absence of Ii, class II molecules have a conformation which suggests that they are empty of peptide (5, 6). We have evaluated the effect of Ii on the presentation of endogenous proteins, using bona

fide antigen-presenting cells (APCs) from mice derived by crossing a line devoid of Ii with transgenic lines expressing a segment of myelin basic protein (MBP) under the control of an MHC class II gene promoter. Processing and presentation of the various epitopes within the transgenic MBP segment was read out with a panel of MBP-reactive T cell hybridomas.

Mutant mice lacking the invariant chain (Ii⁰) have been described previously (5). When crossed onto the H-2^b genetic background, the Ii⁰ mutation leads to somewhat reduced A* surface levels but full ability to present exogenous peptide, as when carried on the H-2^b background (5, 6). Transgenic mice expressing a fragment of MBP (Tg) were produced as part of another study and will be described in detail elsewhere (9). As illustrated in Fig. 1A, they produce a chimeric protein consisting of residues 84 to 105 of MBP [MBP(84–105)] inserted between position 43 and 44 of hen egg lysozyme (HEL), selected as a relatively innocuous carrier. This MBP fragment contains several overlapping epitopes recognizable in the context of A* and characterized extensively in past studies (10) and that can be distinguished with the peptides shown in Fig. 1B. The chimeric proteins are expressed in secreted form with the HEL signal sequence (sLM43) or are membrane-anchored as a result of fusion with the transmembrane and cytoplasmic domains of influenza A virus hemagglutinin (mLM43). Expression is driven by the MHC class II Eα promoter, which directs transcription in the thymus and in

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS et Unité 184 de Biologie Moléculaire de l'INSERM, Institut de Chimie Biologique, 11, rue Humann, 67000 Strasbourg, France.

*Present address: Nuffield Department of Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, United Kingdom.

†Present address: Wellcome/CRC Institute, Tennis Court Road, Cambridge CB2 1QR, United Kingdom.