Molecular characterization of WDCP, a novel fusion partner for the anaplastic lymphoma tyrosine kinase ALK

NORIKO YOKOYAMA and W. TODD MILLER

Department of Physiology and Biophysics, School of Medicine, State University of New York at Stony Brook, Stony Brook, NY 11794, USA

Abstract. Anaplastic lymphoma kinase (ALK) is a member of the receptor tyrosine kinase superfamily. The *ALK* gene is a site of frequent mutation and chromosomal rearrangement in various types of human cancers. A novel chromosomal translocation was recently identified in human colorectal cancer between the *ALK* gene and chromosome 2, open reading frame 44 (*C2orf44*), a gene of unknown function. As a first step in understanding the oncogenic properties of this fusion protein, *C2orf44* cDNA was cloned and the encoded protein was characterized, which was designated as WD repeat and coiled coil containing protein (WDCP). A C-terminal proline-rich segment in WDCP was shown to mediate binding to the Src homology 3 domain of the Src family kinase hematopoietic cell kinase (Hck). Co-expression with Hck lead to tyrosine phosphorylation of WDCP. Chromatographic fractionation of WDCP-containing lysates indicates that the protein exists as an oligomer in mammalian cells. These results suggest that, in the context of the *ALK-C2orf44* gene fusion, WDCP imposes an oligomeric structure on ALK that results in constitutive kinase activation and signaling.

Introduction

Somatically acquired mutations are important to the development of cancer. It has been estimated that 1-2% of all human genes are implicated in cancer via mutation (1). Of these, ~90% have somatic mutations, 20% bear germline mutations that predispose to cancer and 10% show somatic and germline mutations (http://www.sanger.ac.uk/genetics/CGP/Census). In unbiased surveys, protein kinases have emerged as one of the most frequently mutated gene families in cancer (1-3).

Correspondence to: Professor W. Todd Miller, Department of Physiology and Biophysics, School of Medicine, State University of New York at Stony Brook, Basic Science Tower, T-6, Stony Brook, NY 11794, USA
E-mail: todd.miller@stonybrook.edu

Additional genome sequencing efforts in diverse cancers have focused specifically on the protein kinase gene family, leading to a large database of known mutations (4). In numerous cases, the potential functional effects of these mutations are unknown. Thus, it is important to conduct follow-up biochemical studies to identify which mutants may represent the drivers that increase cell proliferation.

A recent study of genomic DNA from colorectal cancer patients identified a novel gene fusion event between the anaplastic lymphoma kinase (ALK) and a previously uncharacterized gene, chromosome 2, open reading frame 44 (*C2orf44*) (5). ALK is a receptor-type tyrosine kinase that has been implicated in several gene fusion events in human cancer (6-10). A chromosomal translocation associated with anaplastic large-cell lymphomas produces a fusion protein between ALK and nucleophosmin (NPM) (10). A fusion between the echinoderm microtubule-associated protein-like 4 (EML4) and ALK was identified in a subset of non-small cell lung cancer (11). The *C2orf44-ALK* fusion results from a large (5 million base pair) tandem duplication and results in ALK kinase overexpression (5). The function of the *C2orf44*-encoded protein in normal or cancer cells is unknown.

Our previous study conducted a proteomic screen to identify signaling components in U937 human monocytic cells that interact with the non-receptor tyrosine kinase (NRTK) hematopoietic cell kinase (Hck) (12). The screen was based on the ability of Hck substrates and binding proteins to interact with the Src homology 3 (SH3) domain of the enzyme, a critical regulatory region. In addition to identifying several regulators of phagocytosis, migration and the actin cytoskeleton, an interaction between Hck and the C2orf44 protein was found (12). In the present study, the C2orf44 protein was cloned and characterized as a first step in understanding its cellular function. Due to the presence of conserved sequence elements, the C2orf44 protein was designated as WD repeat and coiled coil containing protein (WDCP). WDCP was demonstrated to bind tightly and specifically to the SH3 domain of Hck through a proline-rich motif at the C-terminus. WDCP was also shown to exist as an oligomer when expressed in mammalian cells. These results indicate that the *C2orf44-ALK* fusion in colorectal cancer may activate ALK through increased kinase oligomerization and trans-phosphorylation.

Materials and methods

Materials. Antibodies against phosphotyrosine (4G10, cat. no. 05-321) and Hck (cat. no. 06-833) were from EMD Millipore (Billerica, MA, USA). M45 antibody was a gift from Dr Pat Hearing (Stony Brook University, NY, USA). Protein A, buffers, antibiotics and other chemicals were from Sigma-Aldrich (St. Louis, MO, USA). Glutathione agarose was purchased from Molecular Probes/Life Technologies (Carlsbad, CA, USA). The mammalian expression vector for Hck was as described previously (12).

Cloning WDCP. Polymerase chain reaction (PCR) was used to amplify the WDCP coding sequence from human cDNA, clone HEP04702 (obtained from the Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan). The primers used were: forward, 5', GGATCCTGA TGGAGTTGGGAAAAGGAAAACTACTCAGG; and reverse 3', AAGCTTTCAAGCCATGCCATCTACATGGTT ACAACAGCC. The WDCP cDNA was cloned into the *Bam*HI and *Hin*dIII sites of plasmid pCM45 (a gift from Dr Pat Hearing, Stony Brook University) to produce an N-terminal M45-tagged version. Protein features were predicted with the Universal Protein Resource UniProt (13).

Cell culture, transfections, immunoprecipitations and western blotting. Human embryonic kidney (HEK) 293 cells were maintained in Dulbecco's modified Eagle's medium with 10% fetal bovine serum, 100 U/ml streptomycin sulfate and 100 μg/ml amphotericin B. Transfections were carried out with Mirus TransIT (Mirus Corp., Madison, WI, USA) using HEK293 cells, which had been plated 24 h previously. The cells were harvested 48 h after transfection and were lysed in buffer containing 20 mmol/l Tris (pH 8.0), 5 mmol/l EDTA, 1% Nonidet P-40, 150 mmol/l NaCl and 2 mmol/l sodium orthovanadate, with protease inhibitors (5 mg/l aprotinin, 5 mg/l leupeptin and 0.1 mmol/l phenyl-methylsulfonyl fluoride) at 4˚C for 30 min. The cell lysates were centrifuged at 14,000 x g for 15 min at 4˚C. Following protein concentration determination, 1 mg/ml concentrations of the cell lysates were used for immunoprecipitation experiments. The lysates were first precleared with protein A beads for 1 h at 4˚C. The protein A beads were discarded following centrifugation. Subsequently, 1 μg antibody (or the immunoglobulin G control) and 15 μl protein A beads were added to the precleared cell lysates and incubated for 1 h at 4˚C. The beads were washed three times with lysis buffer and immunoprecipitated proteins were resolved by SDS-PAGE. The proteins were transferred to polyvinylidene difluoride membranes and western blotting was carried out with the appropriate antibodies. The proteins were visualized using horseradish peroxidase-conjugated secondary antibody and an enhanced chemiluminescent detection kit (GE Life Sciences, Pittsburgh, PA, USA).

Binding assays. Glutathione S-transferase (GST) or GST-SH3 fusions (Hck, Crk, Grb2 N-terminal SH3 and Nck) were immobilized on glutathione-agarose. HEK293 cell lysates were added to the resins and agitated at 4˚C for 30 min. The resins were washed extensively with lysis buffer and bound proteins were eluted with SDS sample buffer and analyzed by SDS-PAGE and western blotting.

Gel filtration. HEK293 lysates were fractionated on a Superdex-200 fast protein liquid chromatography column (GE Life Sciences, Pittsburgh, PA, USA) pre-equilibrated with 20 mmol/l Tris (pH 8.0), 0.2 M NaCl, 10% glycerol and 0.2 mmol/l Na_3VO_4. The proteins were eluted from the column at a flow rate of 0.15 ml/min. Fractions of 0.15 ml were collected. The column fractions were subjected to anti-M45 immunoprecipitation, and assayed for WDCP using anti-M45 western blotting. The molecular weight calibration was performed using protein standards (Sigma-Aldrich) and by constructing a standard curve of retention time vs. molecular weight.

Results

C2orf44 protein. We previously conducted a proteomic screen for the binding partners of the SH3 domain of the Hck non-receptor tyrosine in the U937 human monocytic cell line. One of the proteins identified was the product of the human *C2orf44* gene (12). PCR was used to amplify the full-length cDNA encoding the C2orf44 protein. The protein consists of 721 amino acids (Fig. 1). The amino-terminal half of the protein contains two WD repeats (WD1 and WD2); these units are involved in a range of biological functions, including apoptosis, transcriptional regulation and signal transduction (14). The carboxy-terminal half of the protein contains a proline-rich sequence (PPRLPQR) that is similar to known ligands for the SH3 domains (15). The C-terminus is also predicted to have a leucine-rich coiled coil region (CC). In view of these sequence elements, the *C2orf44*-encoded protein was designated as WDCP.

Western blotting for WDCP analysis. CC regions are frequently involved in protein oligomerization, often through the formation of amphiphilic structures. WD repeats have also been observed to mediate oligomerization (14). The *C2orf44-ALK* gene fusion in colorectal cancer could potentially be activated by the inappropriate oligomerization of the ALK tyrosine kinase. To test whether WDCP exists as an oligomer, a mammalian expression vector encoding WDCP with an N-terminal M45 tag was produced. The protein was expressed in HEK293 cells and cell lysates were fractionated on a Superdex-200 gel filtration column that had been calibrated with molecular weight standards. WDCP was visualized in column fractions by western blotting with anti-M45 antibodies (Fig. 2). Although the predicted molecular weight of WDCP is 79 kilodaltons (kDa), relatively low amounts of the protein were detected in this mass range. Instead, the majority of the proteins co-eluted in the same region as higher (130-140 and >200 kDa) molecular weight standards. These results are consistent with the formation of WDCP oligomers in mammalian cells.

Subsequently, the ability of WDCP to interact with SH3 domains was tested. M45-tagged WDCP was expressed in HEK293 cells and incubated lysates with the GST-tagged Hck SH3 domain, immobilized on glutathione agarose. In certain reactions, a Pro-rich competitor peptide (YEVPPPVPPRRR, a peptide we previously showed binds tightly to the Hck SH3

A

B

Figure 1. Structure of WD repeat and coiled coil containing protein (WDCP). (A) Schematic representation of the domain structure of WDCP. The sequence of the proline-rich segment is presented. (B) Amino acid sequence of WDCP, with important sequence elements highlighted. WD1 and WD2, WD repeat domains; CC, coiled-coil domain.

Figure 2. WD repeat and coiled coil containing protein (WDCP) exists as an oligomer. M45-tagged WDCP was expressed in human embryonic kidney 293 cells and lysates were fractionated on a Superdex-200 fast protein liquid chromatography column. Individual fractions were analyzed by M45 western blotting. The elution positions of molecular weight standards are shown above the blot.

domain) was included (16). WDCP bound to the immobilized Hck SH3 domain in the absence of the competitor peptide (Fig. 3A). Binding was substantially reduced in the presence of 50 μM Pro-rich peptide and eliminated altogether in the presence of 500 μM peptide (Fig. 3A). WDCP did not bind to control beads containing immobilized GST (Fig. 3A). These experiments validated the results of the proteomic screen and indicated that WDCP binds to the typical ligand-binding site on the Hck SH3 domain, which recognizes proline-rich sequences.

WDCP contains one predicted SH3 ligand sequence between amino acids 539 and 545. This Pro-rich sequence is predicted to be a class 2 motif, with the consensus sequence PXXPXR (15). The two proline residues within this sequence (Pro 540 and Pro 543) were mutated to alanine and the wild-type and mutant forms were expressed in HEK293 cells. Binding experiments were carried out using cell lysates and a variety of immobilized SH3 domains. Wild-type WDCP showed a strong preference for the SH3 domain of Hck; there were trace amounts of binding to the Crk SH3 domain and undetectable binding to the SH3 domains of Grb2 (N-SH3) and Nck (Fig. 3B). The proline to alanine mutations eliminated binding to the SH3 domain of Hck (Fig. 3B). The expression of the Pro-mutant form of WDCP was confirmed to be equally strong as the wild-type (Fig. 3C). Thus, the SH3 binding site was localized on WDCP to the Pro-rich sequence

A
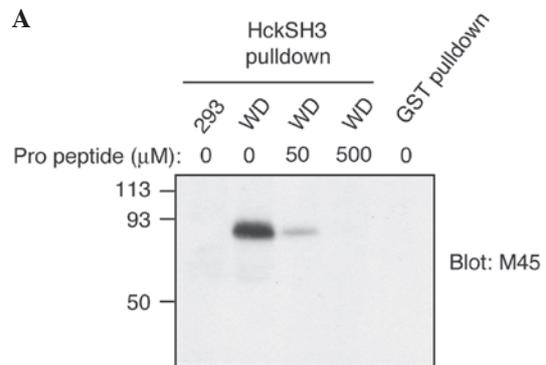
B
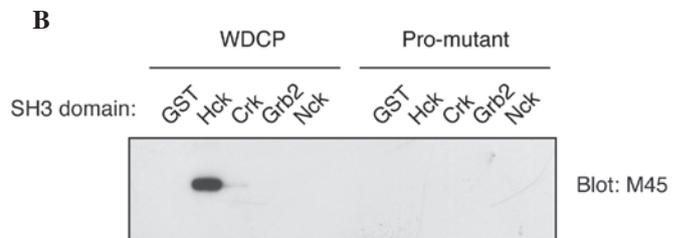
C

Figure 3. Interaction of WD repeat and coiled coil containing protein (WDCP) with Src homology 3 (SH3) domains. (A) Lysates from WDCP-expressing human embryonic kidney (HEK) 293 cells (WD) or untransfected cells (293) were incubated with immobilized hematopoietic cell kinase (Hck) SH3 domain [or glutathione S-transferase (GST) control]. In certain experiments, a proline-rich competitor peptide was added. Following washing, the bound proteins were eluted with SDS-PAGE sample buffer and visualized by anti-M45 western blotting. (B) HEK293 cells expressing WDCP (wild-type or a P540A/P543A mutant) were analyzed in pulldown reactions, as in panel A, using four different immobilized SH3 domains (or GST as a control). (C) Control HEK293 cells or cells expressing WDCP (wild-type or a P540A/P543A mutant) were lysed and analyzed by SDS-PAGE and anti-M45 western blotting. Crk, Grb2, Nck.
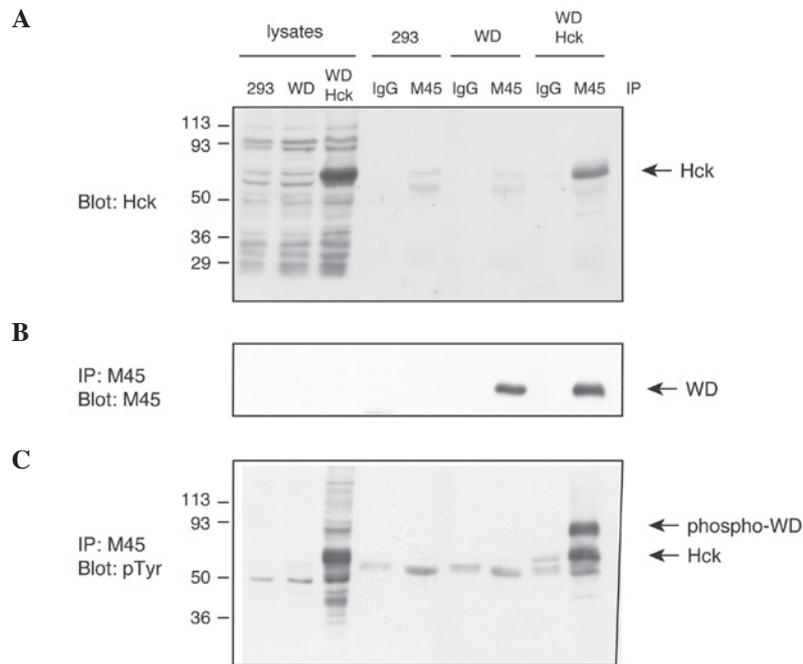
Figure 4. Hematopoietic cell kinase (Hck) binds and phosphorylates WD repeat and coiled coil containing protein (WDCP). (A) Human embryonic kidney (HEK) 293 cells were transfected with M45-tagged WDCP alone (WD) or WDCP plus Hck. In the left three lanes, cell lysates were analyzed directly by SDS-PAGE. In the remaining lanes, anti-M45 [or immunoglobulin G (IgG) control] immunoprecipitation (IP) reactions were carried out. The immunocomplexes were resolved by SDS-PAGE and analyzed by anti-Hck western blotting. The position of Hck is indicated. (B) The membrane was reprobed with anti-M45 antibody to show the presence of WDCP in the immunocomplexes. (C) The membrane was reprobed with anti-phosphotyrosine antibody. The positions of Hck and phosphorylated WDCP are indicated.

between 539-545 and it was demonstrated that WDCP has selectivity with respect to the SH3-mediated protein-protein interactions in which it is involved.

Interaction of WDCP and Hck. Subsequently, the interaction of WDCP with full-length Hck kinase in mammalian cells was tested. M45-tagged WDCP was expressed in HEK293 cells, either alone or together with Hck. Following immunoprecipitation of WDCP, anti-Hck western blotting experiments were performed to show that the two proteins interact (Fig. 4A). The levels of precipitated WDCP were similar in the presence or absence of Hck (Fig. 4B). The blot was stripped and reprobed with anti-phosphotyrosine antibodies (Fig. 4C). These experiments showed that WDCP is tyrosine phosphorylated when co-expressed with Hck, but not when expressed alone. Thus, Hck promotes WDCP tyrosine phosphorylation, either directly (via an enzyme-substrate interaction) or indirectly (for example, by stimulating another kinase that directly phosphorylates WDCP).

Discussion

ALK is a member of the superfamily of RTKs (6,9). The *ALK* gene is localized on chromosome 2p23.1. ALK knockout mice develop normally, with a normal lifespan, exhibiting only behavioral and neurochemical alterations (17). Similar to other RTKs, ALK contains an extracellular ligand binding domain, a single-pass transmembrane sequence and an intracellular tyrosine kinase catalytic domain. Ligand binding to the extracellular portion of ALK triggers trans-autophosphorylation of tyrosine residues within the kinase domain. Once

phosphorylated, the ALK kinase domain is fully active and phosphorylates downstream substrates to promote cell growth and survival. The downstream signaling partners activated by ALK include phospholipase Cγ, mitogen-activated protein kinase, phosphoinositide 3'-kinase/AKT and signal transducer and activator of transcription (STAT) proteins (6,9).

The *ALK* gene is involved in a number of chromosomal rearrangements in human cancers (6,7). Frequently, these events produce fusion proteins containing the intracellular tyrosine kinase domain of ALK; the fusion proteins dimerize to give constitutive kinase activation. These ALK fusions have been described as oncogenic drivers in solid tumors and hematopoietic cancers (7-9). In non-small cell lung cancer, multiple *EML4* breakpoints fuse in frame with exon 20 of *ALK* to produce activated EML4-ALK fusions. These fusions are oncogenic in cell lines (11) and produce lung adenocarcinoma in transgenic mice (18). The Food and Drug Administration-approved ALK small molecule kinase inhibitor crizotinib is a clinically effective therapy for patients harboring the EML4-ALK translocation; a response rate of 57% and a disease control rate of 90% were reported in a group of 82 patients (7-9). The fusion between ALK and NPM in anaplastic large-cell lymphoma was the first chromosomal translocation to be reported involving the *ALK* gene (10). Similar to the situation with the EML4-ALK fusion, this translocation leads to constitutive ALK kinase activity and stimulation of downstream signaling pathways. In addition to chromosomal translocations, somatic and germline point mutations in the kinase or juxtamembrane domains of ALK have been reported in neuroblastoma and anaplastic thyroid cancer (6,9). Whereas the biochemical effects of all these

mutations have not been investigated, they also presumably lead to enhanced/constitutive ALK kinase activity.

In the present study, a novel fusion partner for ALK was characterized: The protein encoded by the *C2orf44* gene, for which we propose to be known as WDCP. This gene fusion was identified in a colorectal cancer patient. The resulting fusion protein contains the N-terminal 680 amino acid residues fused in-frame to ALK residues 1061-1620 (5). This portion of ALK comprises the entire tyrosine kinase catalytic domain and C-terminal region. The attachment of the WDCP residues presumably leads to increased ALK activity by promoting oligomerization, which has been observed for other ALK fusions. *C2orf44* cDNA was cloned and the WDCP protein was expressed in mammalian cells. WDCP was demonstrated to exist as a high-molecular weight protein, supporting this idea. A proline-rich segment within WDCP (residues 539-545, contained in the C2orf44-ALK fusion) was also shown to bind specifically to the SH3 domain of the Src family kinase Hck. Furthermore, WDCP serves as a substrate for Hck. This is reminiscent of the situation in anaplastic large-cell lymphoma, in which the NMP-ALK fusion serves as a Src kinase substrate (19). Furthermore, treatment of anaplastic large-cell lymphoma cells with Src kinase inhibitors or Src-specific small interfering RNA decreases their proliferation rate (19). In conclusion, these results on WDCP indicate that such treatments may also be successful against the C2orf44-ALK fusion.

Acknowledgements

References

1. Futreal PA, Coin L, Marshall M, *et al*: A census of human cancer genes. Nat Rev Cancer 4: 177-183, 2004.
2. Pleasance ED, Cheetham RK, Stephens PJ, *et al*: A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 463: 191-196, 2010.
3. Pleasance ED, Stephens PJ, O'Meara S, *et al*: A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature 463: 184-190, 2010.
4. Greenman C, Stephens P, Smith R, *et al*: Patterns of somatic mutation in human cancer genomes. Nature 446: 153-158, 2007.
5. Lipson D, Capelletti M, Yelensky R, *et al*: Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. Nat Med 18: 382-384, 2012.
6. Chiarle R, Voena C, Ambrogio C, Piva R and Inghirami G: The anaplastic lymphoma kinase in the pathogenesis of cancer. Nat Rev Cancer 8: 11-23, 2008.
7. Shackelford RE, Vora M, Mayhall K and Cotelingam J: ALK-rearrangements and testing methods in non-small cell lung cancer: a review. Genes Cancer 5: 1-14, 2014.
8. Morales La Madrid A, Campbell N, Smith S, Cohn SL and Salgia R: Targeting ALK: a promising strategy for the treatment of non-small cell lung cancer, non-Hodgkin's lymphoma, and neuroblastoma. Target Oncol 7: 199-210, 2012.
9. Hallberg B and Palmer RH: Mechanistic insight into ALK receptor tyrosine kinase in human cancer biology. Nat Rev Cancer 13: 685-700, 2013.
10. Morris SW, Kirstein MN, Valentine MB, *et al*: Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. Science 263: 1281-1284, 1994.
11. Soda M, Choi YL, Enomoto M, *et al*: Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 448: 561-566, 2007.
12. Scott MP, Zappacosta F, Kim EY, Annan RS and Miller WT: Identification of novel SH3 domain ligands for the Src family kinase Hck. Wiskott-Aldrich syndrome protein (WASP), WASP-interacting protein (WIP), and ELMO1. J Biol Chem 277: 28238-28246, 2002.
13. UniProt Consortium: Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res 42 (Database issue): D191-D198, 2014.
14. Smith TF, Gaitatzes C, Saxena K and Neer EJ: The WD repeat: a common architecture for diverse functions. Trends Biochem Sci 24: 181-185, 1999.
15. Mayer BJ and Eck MJ: SH3 domains. Minding your p's and q's. Curr Biol 5: 364-367, 1995.
16. Nguyen JT, Porter M, Amoui M, Miller WT, Zuckermann RN and Lim WA: Improving SH3 domain ligand selectivity using a non-natural scaffold. Chem Biol 7: 463-473, 2000.
17. Bilsland JG, Wheeldon A, Mead A, *et al*: Behavioral and neurochemical alterations in mice deficient in anaplastic lymphoma kinase suggest therapeutic potential for psychiatric indications. Neuropsychopharmacology 33: 685-700, 2008.
18. Soda M, Takada S, Takeuchi K, *et al*: A mouse model for EML4-ALK-positive lung cancer. Proc Natl Acad Sci USA 105: 19893-19897, 2008.
19. Cussac D, Greenland C, Roche S, *et al*: Nucleophosmin-anaplastic lymphoma kinase of anaplastic large-cell lymphoma recruits, activates, and uses pp60c-src to mediate its mitogenicity. Blood 103: 1464-1471, 2004.