

## Primary Structure of the Human *fgr* Proto-Oncogene Product p55<sup>c-fgr</sup>

SHIGERU KATAMINE,<sup>1</sup> VICENTE NOTARIO,<sup>2</sup> C. DURGA RAO,<sup>1</sup> TORU MIKI,<sup>1</sup> MARC S. C. CHEAH,<sup>1</sup>  
STEVEN R. TRONICK,<sup>1</sup> AND KEITH C. ROBBINS<sup>1\*</sup>

Laboratory of Cellular and Molecular Biology, National Cancer Institute, Bethesda, Maryland 20892,<sup>1</sup> and Department of Biochemistry, Georgetown University Medical School, Washington, D.C. 20007<sup>2</sup>

Received 5 August 1987/Accepted 7 October 1987

Normal human *c-fgr* cDNA clones were constructed by using normal peripheral blood mononuclear cell mRNA as a template. Nucleotide sequence analysis of two such clones revealed a 1,587-base-pair-long open reading frame which predicted the primary amino acid sequence of the *c-fgr* translational product. Homology of this protein with the *v-fgr* translational product stretched from codons 128 to 516, where 32 differences among 388 codons were observed. Sequence similarity with human *c-src*, *c-yes*, and *fyn* translational products began at amino acid position 76 of the predicted *c-fgr* protein and extended nearly to its C-terminus. In contrast, the stretch of 75 amino acids at the N-terminus demonstrated a greatly reduced degree of relatedness to these same proteins. To verify the deduced amino acid sequence, antibodies were prepared against peptides representing amino- and carboxy-terminal regions of the predicted *c-fgr* translational product. Both antibodies specifically recognized a 55-kilodalton protein expressed in COS-1 cells transfected with a *c-fgr* cDNA expression plasmid. Moreover, the same protein was immunoprecipitated from an Epstein-Barr virus-infected Burkitt's lymphoma cell line which expressed *c-fgr* mRNA but not in its uninfected *fgr* mRNA-negative counterpart. These findings identified the 55-kilodalton protein as the product of the human *fgr* proto-oncogene.

A large body of evidence has implicated certain genetic events affecting a set of normal cellular genes, termed proto-oncogenes, in the process of malignant transformation (3). Alterations in their structure or regulation are mechanisms by which proto-oncogenes can be activated to become oncogenic (1). Of the proto-oncogenes described to date, approximately half encode protein-tyrosine kinases or have structural similarities with genes specifying such enzymes. Two oncogenic protein-tyrosine kinases, the *v-erbB* and *v-fms* gene products, have been identified as altered versions of receptors for polypeptide growth factors (9, 28). These and other findings have suggested that disruption of pathways by which growth factors control tyrosine phosphorylation and cell growth is an important step in the neoplastic process.

Another group of oncogenic protein-tyrosine kinases includes the products of *v-src* and *v-yes* (15, 30). The complete structure of *src* and *yes* proto-oncogene products is known, but unlike polypeptide growth factor receptors, these proteins lack extracellular ligand-binding and hydrophobic membrane-spanning domains (29, 30). Thus, it appears that the tyrosine kinase family consists of subfamilies which can be distinguished on the basis of possessing or lacking components exposed on the cell surface. The *v-fgr* oncogene of Gardner-Rasheed feline sarcoma virus (GR-FeSV) (24) specifies a protein-tyrosine kinase which is closely related to the protein products of *v-src* and *v-yes*. Nucleotide sequence analysis of GR-FeSV has shown that expression of the *v-fgr* oncogene requires helper virus signals for initiation and termination of protein synthesis (20). Thus, *v-fgr* is truncated at both ends with respect to its normal cellular counterpart. Although portions of the *fgr* proto-oncogene have been described (21, 32), sequences flanking its *v-fgr*-related domain have not yet been identified.

Normal cellular *src* and *yes* genes appear to be expressed

in a wide range of normal and transformed cells (12, 26), whereas expression of *c-fgr* is more limited. To date, transcripts of the *fgr* proto-oncogene have been found only in normal and malignant B lymphocytes infected with Epstein-Barr virus, as well as normal human peripheral blood mononuclear cells (4, 21). In the present study, we sought to isolate cDNA molecules representing *c-fgr* mRNA as an approach to elucidating the primary structure of its translational product.

### MATERIALS AND METHODS

**Cells.** Mononuclear cells were prepared from normal human peripheral blood by Ficoll-Hypaque centrifugation. Established cell lines COS-1 (11), the Burkitt's lymphoma-derived BJAB (16), and Epstein-Barr virus-infected BJAB, designated BJAB-GC (16), were also used.

**Construction of cDNA library.** RNA was extracted from mononuclear cells by the method of Chirgwin et al. (5) and selected twice by passage over an oligo(dT)-cellulose column (Collaborative Research, Waltham, Mass.). The method of Okayama and Berg was used to construct a cDNA library (22). A probe representing *v-fgr* (*pv-fgr-1*) (32) was used for screening cDNA clones.

**Analysis of cellular RNA and DNA.** DNAs were digested with restriction enzymes, fractionated by electrophoresis through agarose gels, and blotted onto nitrocellulose filters as described (32). Cellular RNAs were fractionated in the presence of formaldehyde by agarose gel electrophoresis and transferred to nitrocellulose filters (4). Filters were hybridized with nick-translated DNA fragments, and hybridizing bands were visualized by autoradiography.

**Nucleotide sequence determination.** Restriction enzyme fragments derived from cDNA clones were transferred into the polylinker regions of M13mp18 and 19. Sequencing reaction products were analyzed in buffer gradient or 6% polyacrylamide gels. In each case, the nucleotide sequence for both strands was determined. Sequences were assembled with the aid of the Intelligenetics GEL program.

\* Corresponding author.

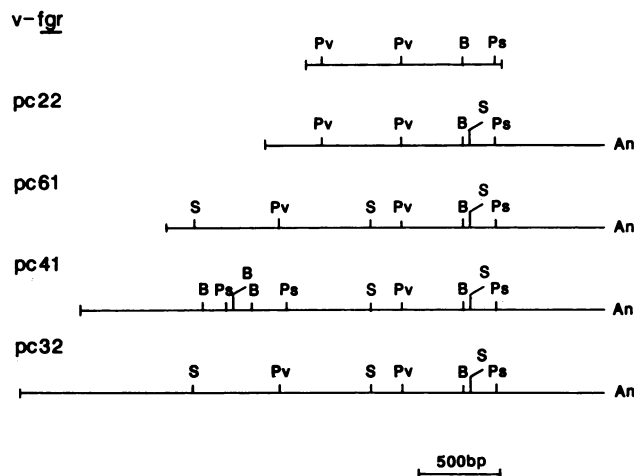


FIG. 1. Physical maps of *c-fgr* cDNA clones. The sites of restriction enzyme cleavage were determined by double-digestion analysis. DNA products were visualized by ethidium bromide staining after agarose gel electrophoresis. Abbreviations: Pv, *PvuII*; B, *BamHI*; Ps, *PstI*; S, *SacI*; An, poly(A) sequence.

**Transfection and protein analysis.** COS-1 cells were plated at a density of  $2 \times 10^6$  per 10-cm petri dish and transfected with plasmid DNAs 24 h later. At 48 h after transfection, cultures were metabolically labeled for 3 h as described (10). BJAB and BJAB-GC cells were grown in suspension to  $5 \times 10^6$ /ml, harvested by centrifugation, and plated in methionine- and cysteine-free Dulbecco modified Eagle medium without serum. Thirty minutes later, cells were metabolically labeled with [ $^{35}$ S]cysteine and [ $^{35}$ S]methionine as described (10). Cultures were lysed with 1 ml of buffer containing 10 mM sodium phosphate, pH 7.5, 100 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate, and 0.1 mM phenylmethylsulfonyl fluoride. Clarified extracts were apportioned into 200- $\mu$ l portions and then incubated with 3  $\mu$ l of antiserum for 60 min at 4°C. Immunoprecipitates were recovered with the aid of protein A bound to Sepharose beads (Pharmacia) and analyzed in 14% polyacrylamide gels by electrophoresis. Synthetic peptides representing residues 16 to 30 (*fgr* N terminus) or 492 to 507 (*fgr* C terminus) were conjugated with thyroglobulin, as described (25), and used as immunogens to raise antisera in rabbits.

## RESULTS

**Isolation of human *c-fgr* cDNA clones.** Previous studies have identified normal human peripheral blood mononuclear cells as a rich source of *fgr* proto-oncogene mRNA (4). Thus, by using polyadenylated [poly(A)<sup>+</sup>] RNA isolated from these cells, cDNA libraries were constructed by the method of Okayama and Berg (22) and screened with a *v-fgr* DNA probe. Ten colonies scored as positive and were further examined by restriction enzyme mapping. As shown in Fig. 1, positive clones contained cDNA inserts of 3.6, 3.2, 2.7, or 2.1 kilobase pairs (kbp), the latter being the most prevalent class. Over a 3' stretch of 1.3 kbp, all four classes of cDNA possessed identical restriction enzyme maps, which were nearly the same as that of *v-fgr* DNA. However, upstream of this region, maps were unique for each class of cDNA insert.

In view of the multiple *src* family members known to be present in human DNA, we first sought to confirm the identity of our cDNA clones. Thus, cDNA inserts were

labeled by nick translation and used as probes to detect homologous sequences present in human DNA by Southern blotting. *v-fgr*, as well as cDNA clones representing each size class, hybridized to a fragment of 2.6 kbp in *BamHI*-treated human DNA. Each cDNA clone detected additional bands of 1.7 and 7.2 kbp (data not shown). Based on the arrangement of *BamHI* sites in the *c-fgr* locus, DNA fragments of 7.2 and 1.7 kbp would lie immediately upstream and downstream, respectively, of the 2.6-kbp fragment hybridized by *v-fgr* probe (21; Tronick and Robbins, unpublished). From these results, we concluded that each of the four cDNAs tested represented the human *c-fgr* locus.

**Nucleotide sequence analysis of cDNA clones.** It has been shown that human *c-fgr* exons 5 and 8 each contain recognition sequences for the restriction endonuclease *PvuII* (21, 32). The most abundant class of cDNA clones, represented by pc22, possessed *PvuII* sites that were separated by approximately 500 bp. By nucleotide sequence analysis of *v-fgr* as well as *c-fgr* exons 4 to 10, the distance between *PvuII* sites should be 489 bp in a cDNA molecule representing the mature *c-fgr* transcript. The location of *PvuII* sites in less-abundant cDNA classes suggested that their structure was unusual. Thus, we initially focused on pc22 and determined its nucleotide sequence. This cDNA was 2,072 bp in length and consisted of a 245-bp stretch upstream and 656 bp downstream of its *v-fgr*-related domain. Only one reading frame was open in the 245-bp upstream region, and this frame was in phase with the *v-fgr* coding sequence. The terminator codon was identified 619 bp upstream of the poly(A) tract. We did not observe the highly conserved eucaryotic polyadenylation signal in the 3' untranslated region but did identify a similar sequence, AAGTAAAA, 26 bp upstream of the poly(A) start site.

No candidate codon for the initiation of protein synthesis was detected in pc22. Thus, to determine whether any of the longer cDNA clones might contain the *c-fgr* amino-terminal coding sequence, an oligonucleotide representing the 5'-most portion of pc22 was synthesized, end-labeled, and used as a probe to screen longer cDNA clones. DNA inserts of pc22 and pc41 hybridized with the oligonucleotide sequence, but pc61 and pc32 cDNA clones did not (data not shown). These findings suggested that pc41 might contain additional coding sequences not found in pc22.

The nucleotide sequence of pc41 cDNA was determined and compared with that of pc22. pc41 contained a 282-bp stretch upstream of the oligonucleotide probe sequence as well as 91 bp which overlapped the 5' region of pc22. Downstream of nucleotide 373, the pc41 sequence was interrupted with respect to pc22 by a 549-bp stretch which contained several stop codons in each reading frame (Fig. 2). The nature of this sequence as well as the location of consensus splice donor and acceptor signals suggested that this sequence represented *c-fgr* intron 2. Furthermore, analysis of corresponding genomic DNA revealed a nucleotide sequence matching that of pc41 from its 5' end through exon 3 (data not shown). From all of these findings, we concluded that pc41 contained *c-fgr* intron 2 and at least a portion of exon 2. Similar comparisons of pc22 and *v-fgr* sequences with those of pc41 revealed that the latter clone also contained intron 7 but lacked exon 5. A summary of the structural features of pc22 and pc41 is shown in Fig. 3.

**Determination and verification of the *c-fgr* coding sequence.** When we examined the sequence of pc41 for evidence of a translation initiation codon, a candidate was identified at nucleotides 148 to 150 (Fig. 4). This codon conformed to Kozak's rules (17) and was in phase with the pc22 coding

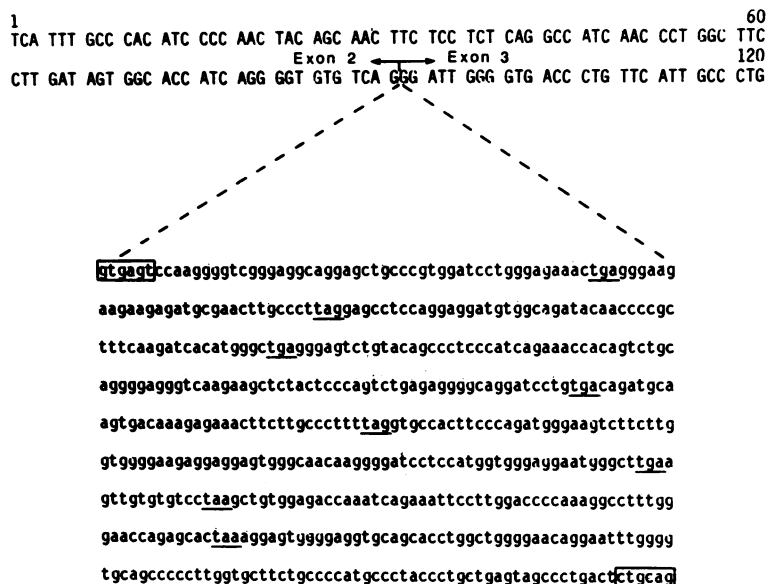


FIG. 2. Location of *c-fgr* intron 2 with respect to sequences of pc22 and pc41. The sequence shown in uppercase letters represents the first 120 nucleotides of pc22. This sequence was also present in pc41 but was interrupted as shown by intron 2, which is shown in lowercase letters. Translation termination codons within intron 2 are underlined. Consensus splice donor (AGgtgag) and acceptor (ctgcag) sequences are boxed at the 5' and 3' extremities, respectively, of the intron sequence.

sequence which began at nucleotide 283. A splice acceptor signal, CTGTAG, was identified at position 129 to 134 and was preceded by stop codons in all three reading frames, indicating that the sequence from position 1 to 134 represented intron 1. Thus, the methionine codon at position 148 to 150 appeared to lie only 13 bp downstream of the intron 1-exon 2 border. An analogous location has been described for the initiator codon of both chicken and human p60<sup>c-src</sup> (30, 31). Alignment of pc41 and pc22 cDNA sequences revealed a long reading frame commencing at nucleotide 148 in pc41 and terminating 1,587 nucleotides downstream within pc22 (Fig. 4). This reading frame would specify a protein of approximately 58 kilodaltons (kDa).

In an effort to verify the open reading frame identified, we recombined pc41 and pc22 by using a *Bsp*MII site located within each of their exon 4 sequences. The resulting construct contained the entire predicted coding domain as well as intron 2, a portion of intron 1, and 3' untranslated sequences. Upstream of the open reading frame, this plasmid, designated pc41/22, contained a simian virus 40 origin of replication and early promoter, elements which would permit overexpression of the coding sequence in COS-1 cells.

Peptide antibodies representing amino- (anti-*fgr*N) and carboxy- (anti-*fgr*C) terminal regions of the predicted *c-fgr* gene product were prepared and used to examine lysates of metabolically labeled COS-1 cells which were transfected with the pc41/22 construct or with pc22, which lacked amino-terminal coding sequences. As shown in Fig. 5A, both antisera immunoprecipitated a protein of 55 kDa from lysates of pc41/22 but not pc22 transfectants. Moreover, preincubation with homologous peptide blocked the ability of the same antibodies to recognize the 55-kDa protein. Although other bands were also detected with anti-*fgr*C, only the 55-kDa protein (p55) was specifically immunoprecipitated with both antisera. We conclude from these findings that the sequence of the *c-fgr* open reading frame is

correct and that the methionine codon at position 148 to 150 is capable of initiating synthesis of p55.

We next attempted to identify the authentic *fgr* translational product in vivo. Thus, BJAB-GC, a human Burkitt's lymphoma cell line known to express *c-fgr* mRNA (4), and its *c-fgr* mRNA-negative counterpart, BJAB, were analyzed by immunoprecipitation. As shown in Fig. 5B, anti-*fgr*C specifically detected p55 in lysates of BJAB-GC but not BJAB cells. Additional bands detected by this antibody in COS-1 cells transfected with pc41/22 were not observed in BJAB-GC cells. Together, these findings establish p55 as the primary translational product of the human *c-fgr* gene.

**Relatedness of p55<sup>c-fgr</sup> to other protein-tyrosine kinases.** Comparison of the p55<sup>c-fgr</sup> amino acid sequence with that specified by *v-fgr* revealed a region of close similarity stretching from position 128 to 516 (Fig. 6). Within this region, the two proteins differed in only 32 of 388 predicted amino acids. Since *v-fgr* was derived from the feline *c-fgr* gene, most if not all of the observed amino acid differences could represent evolutionary changes. The amino-terminal

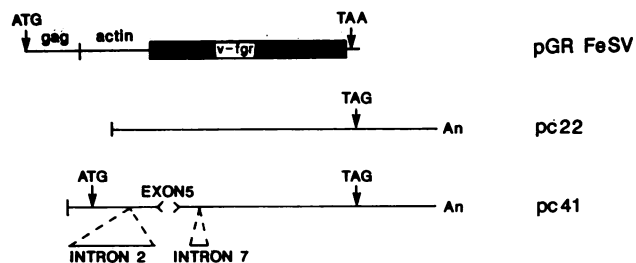


FIG. 3. Organization of pc41 with respect to pc22 and the GR-FeSV transforming gene. The nucleotide sequence of pc41 is aligned with that of pc22 and *v-fgr*. Locations of translation initiator codons (ATG), terminator codons (TAA, TAG), and introns 2 and 7 are indicated. The lack of exon 5 in pc41 is also shown. A<sub>n</sub>, Poly(A) sequence.

GGGAGGACCCCAATCTAGGCCAAAGAGGAAAGGCCAGTGCCTGTATGAGCGTATGAGCATGTGCATGCCGCTGTGTGCACAGGGTGGT  
 GCACCTGGCAGGGGTCCTTGTAGTGAGGCATGCCCCATCTCTAGCAGGGAACCTGGAATGGGCTGTGTGTCTGCAAGAAATGGAGCCG  
 10 20 30 40 50 60 70 80 90  
 100 110 120 130 140 150 160 170 180  
 GTGGCCACGGCCAAAGGAGGATGCTGGCCTGGGAAGGGGACTTCAGAAGCTACGGGGCAGCAGACCCTATGGGCGTGCACCCCACTAAGGCC  
 V A T A K E D A G L E G D F R S Y G A A D H Y G P D P T K A  
 190 200 210 220 230 240 250 260 270  
 CGGCGTGCATCCTATTGCCCCACATCCCAACTACAGCACTTCTCTCAGGCCATCAACCCTGGCTTCTGTATAGTGGCACCATC  
 R P A S S F A H I P N Y S N F S S Q A I N P G F L D S G T I C  
 280 290 300 310 320 330 340 350 360  
 AGGGGTGTGTACGGATTTGGGGTACCTGTTCATTGCCCTGTATGACTATGAGGCTCGAACTGAGGATGACCTCACCTTACCAAGGGC  
 R G V S G I G V T L F I A L Y D Y E A R T E D D L T F T K G  
 370 380 390 400 410 420 430 440 450  
 GAGAAGTTCACATCCTGAACAATACTGAAAGTGACTGGTGGGAGGCTCGTCTCTCAGCTCCGGAAAACTGGCTGCAATCCAGCAAC  
 E K F H I L N N T I E G D W W E A R S L S S G K T G C I P S N  
 460 470 480 490 500 510 520 530 540  
 TACGTGGCCCTGTTGACTCAATCCAAGCTGAAGAGTGGTACTTTGAAAGATTGGGAGAAAGGATGCAGAGAGGCAGCTGCTTTCACCA  
 Y V A P V D S I Q A E E W Y F G K I G R K D A E R Q L L S P  
 550 560 570 580 590 600 610 620 630  
 GGCAACCCAGGGGGCTTCTCATTGGGAAAGCGAGCCACCAAGTGGCTACTCCCTGTCCATCCGGGACTGGGATCAGACCAGA  
 G N P Q G A F L I R E S E T T K G A Y S L S I R D W D Q T R  
 640 650 660 670 680 690 700 710 720  
 GCGCATGTGAAGCATCAAGATCCGCAACTGGACATGGGGCGCTACTACATCACCACACGGGTTCACTTCACTCGGTGCAGGAG  
 G D H V K H Y K I R K L D M G G Y Y T T R V Q F N S V A E  
 730 740 750 760 770 780 790 800 810  
 CTGGTGCAGCACTACATGAGGTGAATGACGGGCTGTGCAACCTGCTCATCGCCCTGCACCATCATGAAGCCGAGAGCCTGGGCGCTG  
 L V Q H Y M E V N D G L C N L L I A P C T I M K P Q T L G L  
 820 830 840 850 860 870 880 890 900  
 GCCAAGGACCGCTGGGAGATCAGCCGACCTCCATCAGCTGGAGCCCGCTGGGACCGGCTGCTTCGGGGATGTGTGGCTGGCCAGC  
 A K D A W E I S R S S I T L E R R L G T G C F G D V W L G T  
 910 920 930 940 950 960 970 980 990  
 TGGAACGGCAGCACTAAGTGGCGGTGAAGACGCTGAAGCCGGGACCATGTCCCGAAGGCCTTCTGGAGGAGGCGCAGGTCAATGAAG  
 W N G S T K V A V K T L K P G T M S P K A F L E A Q V M K  
 1000 1010 1020 1030 1040 1050 1060 1070 1080  
 CTGCTCGCGCAGCACAAGCTGGTGCAGCTGTACGCCGTGGTGTGGAGGAGGCCATCTACATCGTGACCGAGTTCATGTGTACCGCCAGC  
 L L R H D K L V Q L Y A V V S E E P Y I V T E F M C H G S  
 1090 1100 1110 1120 1130 1140 1150 1160 1170  
 TTGCTGGATTTTCTCAAGAACCCAGAGGGCCAGGATTTGAGGCTGCCCAATGGTGGACATGGCAGCCAGGTAGCTAGGGCATGGCC  
 L L D F L K A N P E G Q D L R L P Q L V D M A A Q I V A E G M A  
 1180 1190 1200 1210 1220 1230 1240 1250 1260  
 TACATGGAACGCATGAAGTACATTCACCGCGACCTGAGGGCAGCCAAACATCTGGTGGGGAGCGGCTGGCGTGAAGATCGCAGACTT  
 Y M E R M N Y I H R D L R A A N I L V G E R L A C K I A D F  
 1270 1280 1290 1300 1310 1320 1330 1340 1350  
 GGCTTGGCCGCTCATCAAGGACGATGAGTACAACCCCTGCCAAGTTCCAAGTTCCCATCAAGTGGACAGCCCAAGCTGCCCTC  
 G L A R L I K D E Y N P C Q G S K F P I K W T A P E A A L  
 1360 1370 1380 1390 1400 1410 1420 1430 1440  
 TTTGGCAGATTCCACATCAAGTCAGACGCTGTGGTCTTTGGGATCTGCTCACTGAGCTCATCACCAGGGCCCAATCCCTACCCAGCC  
 F G R F T I K S D V W S F G I L T E L I T K G R I P Y C P G  
 1450 1460 1470 1480 1490 1500 1510 1520 1530  
 ATGAATAAACGGGAAGTGTGGAACAGGTGGAGCAGGGCTACCACATGCCGTGCCCTCCAGGCTGCCAGCATCCCTGTAGGAGGCCATG  
 M N K R E V L E Q V E Q G Y H M P C P P G C P A S L Y E A M  
 1540 1550 1560 1570 1580 1590 1600 1610 1620  
 GAACAGACTGGCGTCTGGACCCGGAGAGGCGCTTCCAGTACCTGTCAGTCTCCTGGAGGACTACTTCACTCCGCTGCAACC  
 E Q T W R L D P E E R P T F E Y L Q S F L E D Y F T S A E P  
 1630 1640 1650 1660 1670 1680 1690 1700 1710  
 CAGTACCACCCGGGGATCAGACATAGCCTGTCCGGGATCAACCCCTCTCTGGCGGTGGCCACCAGTCCCTTGCCAATCCCCAGAGCTGTT  
 Q Y Q P G D Q T  
 1720 1730 1740 1750 1760 1770 1780 1790 1800  
 CTTCCAAAGCCCCAGGCTGGCTTAGAACCCCATAGACTCTAGCATCACCAGGAGCGTGGCTGCTGTGACACCACCTAGGGCAACCTAC  
 1810 1820 1830 1840 1850 1860 1870 1880 1890  
 TTGTTTACAGATGGGGCAAAAGGAGGCCAGAGCTGATCTCTCATCCGCTCTGGCCCAAGCACTATTTCTTCTTTTCCACTTAGGCC  
 1900 1910 1920 1930 1940 1950 1960 1970 1980  
 CCTACATGCGTGTAGCCCTTCTCACTCCACCCAAAAGTGTCTCAGACCTTGTCTAGTTATTTATAAAGTGTATGTACCTCCCTC  
 1990 2000 2010 2020 2030 2040 2050 2060 2070  
 ACTTCTCTCTATCACTGCTTCTCTACTCTCTTTTATCTCACTCTAGTCCAGGTGCCAAGAATTTCCCTTCTACCCCTCTATTCTTGT  
 2080 2090 2100 2110 2120 2130 2140 2150 2160  
 GTCTGTAAAGTTACAAGATCAGAAAAGTCTTGGCTGGACCCCTTCTGCTGGGTGGATGCAGTGGTCCAGGACTGGGGTCTGGGCCAG  
 2170 2180 2190 2200 2210 2220 2230 2240 2250  
 GTTTGAGGAGAAGTTGCAGAGCACTTCCACCTCTCTGAATAGTGTGTGATTATTGATTCTGTAATAAGTAAAATGACA  
 2260 2270 2280 2290 2300 2310 2320 2330 2340  
 ATATGAATCTCCA  
 2350

FIG. 4. Composite nucleotide sequence of *c-fgr* cDNA. The amino acid sequence predicted from the long open reading frame is shown by the single-letter designations. Nucleotide positions are shown above and amino acid positions below their respective sequences. Intron-exon (int-ex) borders are indicated. The nucleotide sequences 283 to 2354 and 1 to 575 were obtained from pc22 and pc41 cDNA, respectively. Intron 2 sequences present in pc41 have been removed. The figure was prepared with the aid of a computer program (27).

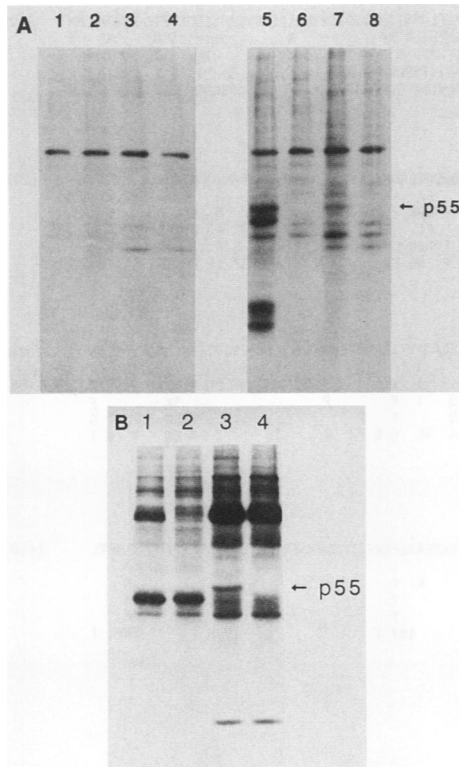


FIG. 5. Identification of the human *c-fgr* translational product. COS-1 cells transfected with pc22 (A, lanes 1 to 4) or pc41/22 (A, lanes 5 to 8) or BJAB (B, lanes 1 and 2) and BJAB-GC (B, lanes 3 and 4) cells were metabolically labeled with [ $^{35}$ S]methionine, lysed, and clarified. Extracts were immunoprecipitated with anti-*fgr*C (A, lanes 1, 2, 5, and 6; B, lanes 1 to 4) or anti-*fgr*N (A, lanes 3, 4, 7, and 8). In some cases (A, lanes 2, 4, 6, and 8; B, lanes 2 and 4), antisera were preincubated with homologous peptide. Immunoprecipitates were analyzed by polyacrylamide gel electrophoresis as described in Materials and Methods. Location of p55<sup>*c-fgr*</sup> is indicated.

region of p55<sup>*c-fgr*</sup> was not present in GR-FeSV but was replaced by portions of feline leukemia virus *gag* and cellular actin genes. At the carboxy terminus of the viral protein, 13 amino acid residues were replaced with a new set of six amino acids previously shown to derive from the GR-FeSV *env* gene (20). As a result of these changes, Tyr-523, a site analogous to Tyr-527 of p60<sup>*c-src*</sup>, was absent from the *v-fgr* translational product.

The predicted *c-fgr* gene product was highly related to other protein-tyrosine kinases, beginning at position 76 and extending to the terminator codon at position 530 (Fig. 6). From the location of intron 2 in pc41, this region would represent *c-fgr* exons 3 to 12 (21). Cellular *yes*, *fyn*, and *src* gene products were 77, 76, and 74% identical, respectively, to p55<sup>*c-fgr*</sup> from amino acid position 76 onward. The murine *lck* translational product was more distantly related to the *c-fgr* protein within the same region, showing 57% relatedness. Protein-tyrosine kinases, including *abl*, *fes*, and *fms* proto-oncogene products, as well as receptors for epidermal growth factor or platelet-derived growth factor, showed a further reduced degree of similarity to the *c-fgr*-encoded protein.

Exon 2 of the *c-fgr* gene specified the amino-terminal region of p55<sup>*c-fgr*</sup>. Optimal alignment of this region with amino-terminal domains of human *fyn*, *yes*, and *src* proteins showed 31, 11, and 11% identity, respectively (Fig. 6). The only common feature among these four molecules was the

presence of a glycine residue at position 2, a site known to be important for posttranslational myristylation of p60<sup>*c-src*</sup> (23).

**Poly(A)<sup>+</sup> *c-fgr* RNAs detected in peripheral blood mononuclear cells contain introns.** Three of the 10 *c-fgr* cDNA clones examined contained introns. Partial nucleotide sequence analysis of pc61 and pc32 revealed that these clones contained portions of intron 4 at their 5' ends (data not shown). Moreover, pc61, pc41, and pc32 all contained intron 7, and pc41 contained intron 2 as well. Since our longest cDNA clones contained introns, we examined exon sequences upstream of pc22 for regions known to contribute to RNA secondary structure. However, a computer analysis revealed no stable structures which could account for the lack of longer intronless cDNAs in our library. There were multiple translation stop codons in each reading frame within introns 2 and 4, whereas intron 7 contained a single terminator codon which was in frame with the translated sequence. Thus, RNAs containing these introns could not direct the synthesis of p55<sup>*c-fgr*</sup>. However, detection of p55<sup>*c-fgr*</sup> in COS-1 cells transfected with pc41/22 confirmed that intron 2 could indeed be properly excised *in vivo* to permit translation of this protein.

From the isolation frequency of cDNA clones containing introns, we estimated that approximately 30% of steady-state, poly(A)<sup>+</sup> *c-fgr* RNA would be incompletely processed. To directly test this possibility, poly(A)<sup>+</sup> RNA prepared from normal human peripheral blood mononuclear cells was examined by Northern (RNA) blot hybridization with *v-fgr* or the pc41 intron 2 sequence as the probe. As shown in Fig. 7, the *v-fgr* probe detected a band of 3.6 kb as well as the major 3.0-kb mRNA. In contrast, the intron 2 sequence probe detected 3.6-kb but not 3.0-kb poly(A)-containing RNA species. Thus, incompletely processed RNAs are present in mononuclear cells as discrete species and account for the minor 3.6-kb RNA observed. This finding demonstrated that our cDNA clones faithfully represented the *c-fgr* poly(A)<sup>+</sup> RNAs present in normal human peripheral blood mononuclear cells.

## DISCUSSION

In the present study, we have described overlapping cDNA clones representing transcripts of the human *fgr* proto-oncogene. The long open reading frame present in these clones predicted the primary structure of the *c-fgr* translational product and provided an approach to identify this protein *in vivo*. Antibodies designed to recognize terminal regions of the *c-fgr* translational product immunoprecipitated a 55-kDa protein synthesized in COS-1 cells at the direction of the predicted *c-fgr* open reading frame. Detection of the same protein in a human Burkitt's lymphoma cell line known to express *c-fgr* mRNA established p55<sup>*c-fgr*</sup> as the primary translational product of the authentic *c-fgr* gene. The major feature distinguishing this protein from closely related products of *src*, *yes*, and *fyn* genes is the 75-residue polypeptide specified by *c-fgr* exon 2. A similar conclusion based on an identical predicted amino acid sequence for this domain has recently been reported (13). Although cellular roles have not yet been established for this group of proteins, structural evidence would suggest that their functions are similar and that their amino-terminal domains are responsible for unique interactions with cellular targets.

Previous studies showing that human *c-fgr* exons 4 to 10 are identical in size and closely related in sequence to those of human *c-src* have provided strong evidence that *src* and *fgr* genes arose from a common ancestor (21). Exons 2 and 3

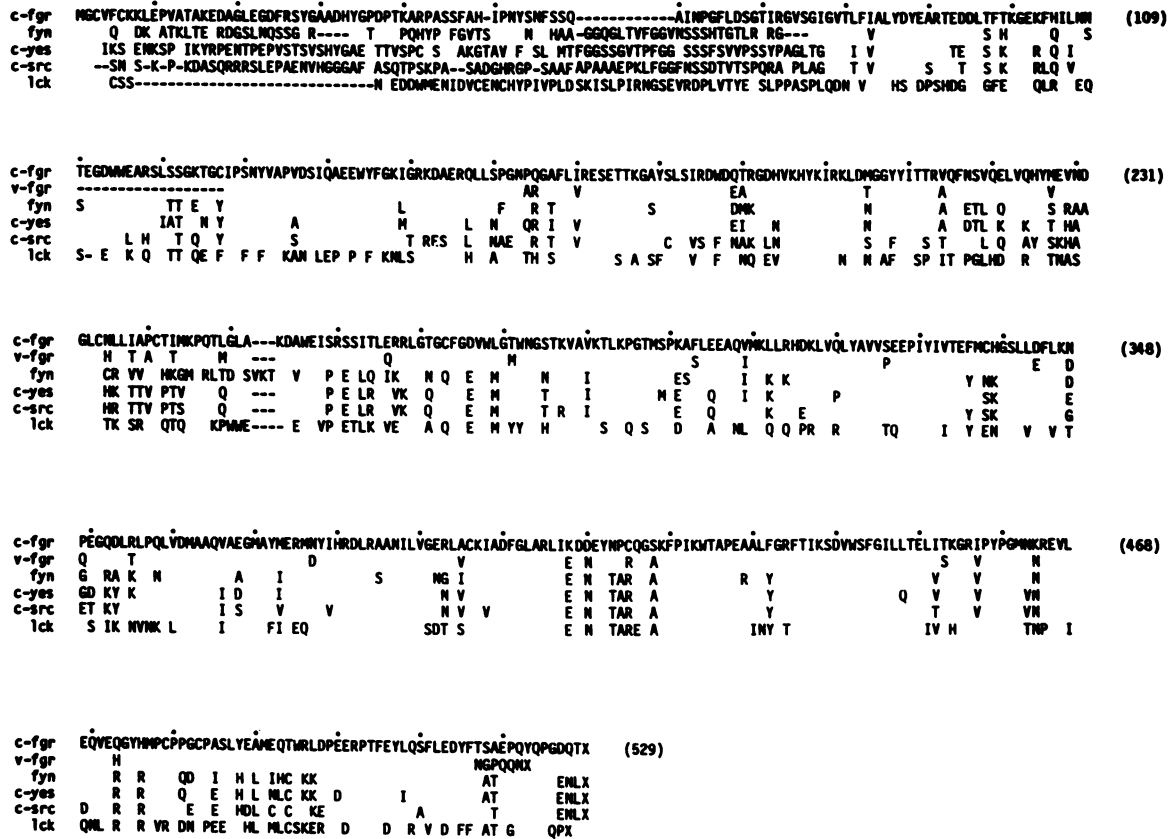


FIG. 6. Comparison of p55<sup>c-fgr</sup> with other *src* gene family translational products. Human *fyn* (14), *c-yes* (29), and *c-src* (31) and mouse *lck* (19, 33) translational products are shown. Dashes indicate gaps generated for optimal alignment. Only amino acid residues differing from those of p55<sup>c-fgr</sup> are shown. X indicates translational stop codons. Sequences were aligned by using the FASTP program of Lipman and Pearson (18).

of human *c-src* have recently been characterized and shown to be 250 and 100 bp in length, respectively (31). In the present study, we have established the extent of *c-fgr* exon 2 as 239 bp and exon 3 as 104 bp, sizes closely approximating their *c-src* counterparts. In addition, polypeptides specified by *c-src* and *c-fgr* exons 3 are highly related. Despite differences in their amino-terminal domains, initiator codons for both p60<sup>c-src</sup> and p55<sup>c-fgr</sup> lie very close to the 5' border of their respective exons 2. These similarities in the organization of human *src* and *fgr* genes suggest that their divergent exons 2, as well as their conserved exons 3 to 12, arose from a common ancestor. If so, the primordial exon 2 has undergone more rapid evolutionary change than exons 3 to 12. Thus, a mechanism involving variable evolutionary rates among exons, not exon shuffling, may account for the differences in the amino-terminal domains of the *src* and *fgr* translational products.

A minor *c-fgr* transcript, 3.6 kb in length, in addition to the most abundant 3.0-kb mRNA, has been observed in normal peripheral blood mononuclear cells as well as certain lymphoid cell lines (4, 21). Multiple transcripts have also been described for another tyrosine kinase gene, the *abl* proto-oncogene (2). In the case of *abl*, these RNA species have been shown to result from alternative 5' exon utilization (2). Furthermore, intron-containing cDNAs have been reported for *lyn*, another new member of the *src* subfamily of genes (34). In the present study, we have shown that the larger *fgr* RNA contains intron 2 and is incompletely processed. Since intron 2 is 549 bp in length, retention of this sequence alone

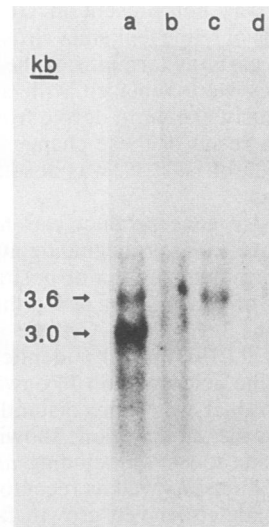


FIG. 7. Detection of intron-containing *c-fgr* RNAs in normal human peripheral blood mononuclear cells. RNA isolated from mononuclear cells was fractionated by oligo(dT)-cellulose chromatography. Poly(A)<sup>+</sup> (lanes a and c) or nonselected flowthrough (lanes b and d) fractions were analyzed by the Northern technique with *v-fgr* (lanes a and b) or *c-fgr* intron 2 (lanes c and d) as the probes. Transcript sizes are shown (in kilobases).

could account for the difference in size of the two *fgr* transcripts. However, our cDNA clones also contained other introns, and intron 7 was common to all cDNAs derived from incompletely processed RNA. Thus, we suspect that the 3.6-kb RNA contains intron 7 as well as intron 2. Recent confirmation that normal granulocytes and monocytes express *fgr* RNA containing intron 7 has been obtained by S1 nuclease protection studies (T. Ley, N. Connally, S. Katamine, R. Senior, and K. C. Robbins, manuscript in preparation). Although we do not yet know whether incompletely processed *fgr* transcripts leave the nucleus, both introns contain stop codons in phase with the translated sequence and would prevent synthesis of the *c-fgr* kinase domain. Thus, RNAs containing intron 2 or 7 could not direct the synthesis of p55<sup>*c-fgr*</sup>. All of these findings suggest that RNA processing may play a role in regulating the expression of p55<sup>*c-fgr*</sup>.

The transforming gene product of GR-FeSV.P70<sup>*gag-actin-fgr*</sup>, consists of retrovirus *gag* and cellular actin sequences in addition to a *v-fgr*-specified protein-tyrosine kinase domain (20). As a result of the multiple recombination events leading to GR-FeSV, the *v-fgr* gene is truncated at both ends with respect to the *c-fgr* coding sequence. Thus, there are major differences between the *fgr* oncogene and proto-oncogene products. At the amino terminus, 127 amino acids were deleted from the *c-fgr* protein and replaced by 118 *gag* and 152 actin residues. The last six amino acids of P70<sup>*gag-actin-fgr*</sup> were derived from the feline leukemia virus envelope gene and replaced a 13-amino-acid stretch present in the *fgr* proto-oncogene product. More subtle differences exist between p60<sup>*c-src*</sup> and its transforming version, p60<sup>*v-src*</sup>. The normal protein contains a tyrosine residue at position 527 which plays a role in regulating p60<sup>*c-src*</sup> kinase activity (6-8). Because of a deletion-substitution event, this residue is not present in p60<sup>*v-src*</sup> (30). The lack of analogous residues in *v-fgr* and *v-yes* translational products but not their cellular cognates has been suggested as a factor contributing to the transforming activity of *v-fgr* and *v-yes* (7). The availability of cDNA clones representing the entire *c-fgr* coding sequence should make it possible to determine whether this or other genetic alterations can convert the *c-fgr* gene to an oncogene.

#### ACKNOWLEDGMENTS

We thank Claire Pennington, Carole Lengel, and Edward Stephens for assistance with various aspects of this work and Stuart Aaronson for continued support.

S.K. is supported by a Public Health Service International Research Fellowship (F05 TWO3839-01). T.M. is supported by a Japanese Overseas Cancer Fellowship from the Foundation for Promotion of Cancer Research.

#### LITERATURE CITED

- Aaronson, S. A., and S. R. Tronick. 1985. The role of oncogenes in human neoplasia, p. 3-15. *In* V. T. Devita, S. Hellman, and S. A. Rosenberg (ed.), *Important advances in oncology*. J. B. Lippincott Co., Philadelphia.
- Ben-Neriah, Y., A. Bernardis, M. Paskind, G. Q. Daley, and D. Baltimore. 1986. Alternative 5' exons in *c-abl* mRNA. *Cell* **44**: 577-586.
- Bishop, J. M. 1987. The molecular genetics of cancer. *Science* **235**:305-311.
- Cheah, M. S. C., T. J. Ley, S. R. Tronick, and K. C. Robbins. 1986. *fgr* proto-oncogene mRNA induced in B-lymphocytes by Epstein-Barr virus infection. *Nature (London)* **319**:238-240.
- Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**:5294-5299.
- Cooper, J. A., and C. S. King. 1986. Dephosphorylation or antibody binding to the carboxy terminus stimulates pp60<sup>*c-src*</sup>. *Mol. Cell. Biol.* **6**:4467-4477.
- Cooper, J. A., K. L. Gould, C. A. Cartwright, and T. Hunter. 1986. Tyr<sup>527</sup> is phosphorylated in pp60<sup>*c-src*</sup>: implications for regulation. *Science* **231**:1431-1434.
- Courtneidge, S. A. 1985. Activation of pp60<sup>*c-src*</sup> kinase by middle T binding or by dephosphorylation. *EMBO J.* **4**:1417-1477.
- Downward, J., Y. Yarden, E. Mayes, G. Scrase, N. Totty, P. Stockwell, A. Ullrich, and J. Schlessinger. 1984. Close similarity of epidermal growth factor receptor and *v-erbB* oncogene protein sequences. *Nature (London)* **307**:521-527.
- Giese, N. A., K. C. Robbins, and S. A. Aaronson. 1987. The role of individual cysteine residues in the structure and function of the *v-sis* gene product. *Science* **236**:1315-1318.
- Gluzman, Y. 1981. SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* **23**:175-182.
- Hunter, T., and J. A. Cooper. 1985. Protein-tyrosine kinases. *Annu. Rev. Biochem.* **116**:188-204.
- Inoue, K., S. Ikawa, K. Semba, J. Sukegawa, T. Yamamoto, and K. Toyoshima. 1987. Isolation and sequencing of cDNA clones homologous to the *v-fgr* oncogene from a human B lymphocyte cell line, IM-9. *Oncogene* **1**:301-304.
- Kawakami, T., C. Y. Pennington, and K. C. Robbins. 1986. Isolation and oncogenic potential of a novel *src*-like gene. *Mol. Cell. Biol.* **6**:4195-4201.
- Kitamura, N., A. Kitamura, K. Toyoshima, Y. Hirayama, and M. Yoshida. 1982. Avian sarcoma virus Y73 genome sequence and structural similarity of its transforming gene product to that of Rous sarcoma virus. *Nature (London)* **297**:205-208.
- Klein, G., B. Giovanella, A. Westman, J. S. Stehlin, and D. Mumford. 1975. An EBV genome negative cell line established from an American Burkitt lymphoma: receptor characteristics, EBV infectibility and permanent conversion into EBV-positive sublines by *in vitro* infection. *Intervirology* **5**:319-334.
- Kozak, M. 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* **12**:857-872.
- Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science* **227**:1435-1441.
- Marth, J. D., R. Peet, E. G. Krebs, and R. M. Perlmutter. 1985. A lymphocyte-specific protein tyrosine kinase gene is rearranged and overexpressed in the murine T-cell lymphoma LSTRA. *Cell* **43**:393-404.
- Naharro, G., K. C. Robbins, and E. P. Reddy. 1984. Gene product of *v-fgr onc*: hybrid protein containing a portion of actin and a tyrosine-specific protein kinase. *Science* **223**:63-66.
- Nishizawa, M., K. Semba, M. C. Yoshida, T. Yamamoto, M. Sasaki, and K. Toyoshima. 1986. Structure, expression, and chromosomal locations of the human *c-fgr* gene. *Mol. Cell. Biol.* **6**:511-517.
- Okayama, H., and P. Berg. 1983. A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells. *Mol. Cell. Biol.* **3**:280-289.
- Pellman, D., E. A. Garber, F. R. Gross, and H. Hanafusa. 1985. Fine structure mapping of a critical NH<sub>2</sub>-terminal region of p60<sup>*src*</sup>. *Proc. Natl. Acad. Sci. USA* **82**:1623-1627.
- Rasheed, S., M. Barbacid, S. A. Aaronson, and M. B. Gardner. 1982. Origin and biological properties of a new feline sarcoma virus. *Virology* **117**:238-244.
- Robbins, K. C., S. G. Devare, E. P. Reddy, and S. A. Aaronson. 1982. *In vivo* identification of the transforming gene product of simian sarcoma virus. *Science* **218**:1131-1133.
- Semba, K., Y. Yamanashi, M. Nishizawa, J. Sukegawa, M. Yoshida, M. Sasaki, T. Yamamoto, and K. Toyoshima. 1985. Location of the *c-yes* gene on the human chromosome and its expression in various tissues. *Science* **227**:1038-1040.
- Shapiro, M. B., and P. Senapathy. 1986. Automated preparation of DNA sequences for publication. *Nucleic Acids Res.* **14**: 65-73.

28. Sherr, C. J., C. W. Rettenmier, R. Sacca, M. F. Roussel, T. A. Look, and E. Stanley. 1985. The *c-fms* proto-oncogene product is related to the receptor for the mononuclear phagocyte growth factor, CSF-1. *Cell* **41**:665-676.
29. Sukegawa, J., K. Semba, Y. Yamanashi, M. Nishizawa, N. Miyasima, T. Yamamoto, and K. Toyoshima. 1987. Characterization of cDNA clones for the human *c-yes* gene. *Mol. Cell. Biol.* **7**:41-47.
30. Takeya, T., and H. Hanafusa. 1983. Structure and sequence of the cellular gene homologous to the RSV *src* gene and mechanism for generating the transforming virus. *Cell* **32**:881-890.
31. Tanaka, A., C. P. Gibbs, R. R. Arthur, S. K. Anderson, H.-J. Kung, and D. J. Fujita. 1987. DNA sequence encoding the amino-terminal region of the human *c-src* protein: implications of sequence divergence among *src*-type kinase oncogenes. *Mol. Cell. Biol.* **7**:1978-1983.
32. Tronick, S. R., N. C. Popescu, M. S. C. Cheah, D. C. Swan, S. C. Amsbaugh, C. R. Lengel, J. A. DiPaolo, and K. C. Robbins. 1985. Isolation and chromosomal localization of the human *fgr* proto-oncogene, a distinct member of the tyrosine kinase gene family. *Proc. Natl. Acad. Sci. USA* **82**:6595-6599.
33. Voronova, A. F., and B. M. Sefton. 1986. Expression of a new tyrosine protein kinase is stimulated by retrovirus insertion. *Nature (London)* **319**:682-685.
34. Yamanashi, Y., S.-H. Fukushige, K. Semba, J. Sukegawa, N. Miyajima, K.-I. Matsubara, T. Yamamoto, and K. Toyoshima. 1987. The *yes*-related cellular gene *lyn* encodes a possible tyrosine kinase similar to p56<sup>lck</sup>. *Mol. Cell. Biol.* **7**:237-243.