PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :		(11) International Publication Number:
C07K 13/00, 15/28, C12N 15/62	A1	WO 92/13001
		(43) International Publication Date: 6 August 1992 (06.08.92)

(21) International Application Number: PCT/US92/00434

(22) International Filing Date: 17 January 1992 (17.01.92)

(30) Priority data:
643,237 18 January 1991 (18.01.91) US

(71) Applicant: NEW YORK UNIVERSITY [US/US]; 550 First Avenue, Room MSB 153, New York, NY 10016 (US).

(72) Inventors: SCHLESSINGER, Joseph ; 37 Washington Square, West, New York, NY 10011 (US). SKOLNIK, Edward, Y. ; 201 West 89th Street, Apt. 5F, New York, NY 10024 (US). MARGOLIS, Benjamin, L. ; 401 East 34th Street, Apt. N-17E, New York, NY 10016 (US).

(74) Agent: BROWDY, Roger, L.; Browdy and Neimark, 419 Seventh Street, Suite 300, N.W., Washington, DC 20004 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), MC (European patent), NL (European patent), SE (European patent).

Published
With international search report.

(54) Title: A RECEPTOR TYROSINE KINASE TARGET PROTEIN cDNA CLONING METHOD AND hGRB PROTEINS

(57) Abstract

A novel expression cloning method for detection and identification of proteins capable of binding to tyrosine-phosphorylated domains of receptor tyrosine kinases is based on the use of a novel probe. This probe comprises an amino acid sequence derived from the tyrosine-phosphorylated portion of the receptor molecule, or a functional derivative thereof, and has at least one phosphorylated tyrosine residue, lacks the tyrosine kinase portion of the receptor, and is detectably labeled. Also disclosed are a method for preparing the probe, a method for mapping to a chromosome a gene encoding a protein capable of binding to tyrosine-phosphorylated domains of receptor tyrosine kinases, and a method for purifying such a protein with the probe. Two proteins discovered using the above cloning method, GRB-1 and GRB-2, DNA encoding these proteins, and methods for detecting these proteins are also disclosed.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CS	Czechoslovakia
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
MC	Monaco
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NL	Netherlands
NO	Norway
PL	Poland
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SN	Senegal
SU	Soviet Union
TD	Chad
TG	Togo
US	United States of America

A RECEPTOR TYROSINE KINASE TARGET PROTEIN cDNA CLONING METHOD AND hGRB PROTEINS

BACKGROUND OF THE INVENTION

Field of the Invention

The invention in the field of molecular and cell biology relates to a novel method, based on direct expression cloning, for identifying cellular proteins capable of binding to the carboxy-terminal domain of receptor tyrosine kinases and serving as substrates for this enzyme in major signal transduction pathways. The invention also relates to novel proteins identified using this method.

Description of the Background Art

A variety of polypeptide growth factors and hormones mediate their cellular effects by interacting with cell surface receptors having tyrosine kinase enzymatic activity (for review, see Williams, L.T. et al., Science 243:1564-1570 (1989); Ullrich, A. et al., Cell 61:203-212 (1990); Carpenter, G. et al. J. Biol. Chem. 265: 7709-7712 (1990)). The interaction of these ligands with their receptors induces a series of events which include receptor dimerization and stimulation of protein tyrosine kinase activity. For the epidermal growth factor receptor (EGFR) as well as other receptors with tyrosine kinase activity, such as the platelet-derived growth factor receptor (PDGFR), kinase activation and receptor autophosphorylation result in the physical association of the receptor with several cytoplasmic substrates (Ullrich et al., supra).

Two substrates for the EGFR kinase have now been definitively identified in living cells: (a) the phosphatidylinositol specific phospholipase C-gamma (PLC-gamma) and (b) the GTPase activating protein (GAP), a protein which may be in the effector loop of the ras protein (Margolis, B. et al. Cell 57: 1101-1107 (1989b); Meisenhelder, J. et al. Cell 57: 1109-1122 (1989); Molloy, C.J. et al.

Nature 342: 711-714 (1989); Wahl, M.I. et al. J. Biol. Chem. 265: 3944-3948 (1990); Ellis, C. et al. Nature 343: 377-381 (1990); Kaplan, D.R. et al. Cell 61 121-133 (1990)).

Similarly, activated PDGFR was shown to tyrosine phosphorylate, and to become associated with PLC-gamma, GAP, and cellular tyrosine kinases such as $pp60^{src}$ (Gould, K.L. et al., Molec. Cell. Biol. 8:3345-3356 (1988); Meisenhelder, J. et al., Cell 57:1109-1122 (1989); Molloy, C.J. et al., Nature 342:711-714 (1989); Kaplan, D.R. et al., Cell 61:121-133 (1990); Kazlauskas, A. et al., Science 247:1578-1581 (1990); Krypta, R.M. et al., Cell 62:481-492 (1990); Margolis, B. et al., Science 248:607-610 (1990)). While the exact sites responsible for the association of EGFR with either PLC-gamma or GAP have not been completely clarified, recent work has begun to identify regions on both the substrate and receptor which contribute to the association.

SH2 (src homology 2) domains appear to be the regions responsible for the association of several tyrosine kinase substrates with activated growth factor receptors. SH2 domains are conserved sequences of about 100 amino acids found in cytoplasmic non-receptor tyrosine kinases such as pp60src, PLC-gamma, GAP and v-crk (Mayer, B.J. et al., Nature 332:272-275 (1988); Pawson, T. Oncogene 3:491-495 (1988)). While having distinct catalytic domains, all these molecules share conserved SH2 and SH3 (src homology 3) domains and the ability to associate with receptors with tyrosine kinase activity (Anderson, D. et al., Science 250:979-982 (1990)).

Tyrosine kinase activation and receptor autophosphorylation are prerequisites for the association between growth factor receptors and SH2 domain-containing proteins (Margolis, B. et al., Mol. Cell. Biol. 10:435-441 (1990);
Kumjian et al., Proc. Natl. Acad. Sci. USA 86:8232-8239 (1989); Kazlauskas, A. et al., Science 247:1578-1581 (1990)). In particular, the carboxy-terminal (C-terminal) fragment of the EGFR, which contains all the known autophosphorylation sites, binds specifically to the SH2 domains of GAP and PLC-gamma (see below). Hence, a major site of association exists

between the SH2 domain of these substrate proteins and the tyrosine phosphorylated C-terminal tail of the EGFR.

With the recognition that binding to the activated tyrosine kinase receptor is conserved among several substrate proteins, efforts to identify additional substrates which share these properties have been undertaken. Target proteins which bind to activated receptors have been identified by analysis of proteins that co-immunoprecipitate with growth factor receptors, or that bind to receptors attached to immobilized matrices (Morrison, D.K. et al., Cell 58:649-657 (1989); Kazlauskas, A. et al., EMBO J. 9:3279-3286 (1990)). While the identity of some of these proteins is known, several others detected utilizing these approaches have not been fully characterized. Moreover, it is possible that rare target molecules which interact with activated receptors have not been detected due to the limited sensitivity of these techniques; the actual stoichiometry of binding may be low, and the detergent solution necessary to solubilize proteins may disrupt binding.

Conventional approaches to isolate and clone these proteins have been arduous, requiring the use of large quantities of tissue or cells lines to purify sufficient amounts of protein for microsequence analysis and subsequent conventional cDNA cloning. Therefore, a need for new approaches for the cloning and subsequent isolation and identification of e these proteins is recognized in the art.

SUMMARY OF THE INVENTION

One of the most pressing needs in our effort to understand and gain control over the regulation of cell growth and oncogenesis is having the ability to identify target proteins for tyrosine kinases.

The present inventors have developed a novel expression/cloning system for the rapid cloning of target proteins for tyrosine kinase-type cellular receptors. The cloning method is based on the ability of a certain class of substrates to bind specifically to the tyrosine-phosphorylated carboxy-terminus (C-terminus) of the epidermal growth factor

receptor (EGFR) (see Example VI, below; Margolis, B. et al., EMBO J. 9:4375-4380 (1990)).

The approach conceived by the present inventors has important advantages over conventional cloning methods, including avoidance of the laborious and costly task of purifying potential target proteins for microsequencing analysis. Moreover, the approach of the present invention provides a method for identifying rare target molecules whose association with activation receptors could not otherwise be detected using conventional techniques. Furthermore, this method allows the identification of structurally or functionally related proteins which, though only weakly homologous at the DNA level, are similar in their property of binding to activated receptors with tyrosine kinase activity. The latter ability is important since conventional screening methods used to identify related genes are typically based on low stringency nucleic acid hybridization. Indeed such hybridization-based screening would not have been successful in cloning and identifying the new proteins of the present invention, GRB-1 and GRB-2, because of their lack of similarity at the DNA level.

The present inventors' approach has its roots in methods previously utilized to clone DNA-binding and jun-binding proteins (Singh, H. et al., Cell 52:415-423 (1988); MacGregor, P.F. et al., Oncogene 5:451-458 (1990)). The methods of the present invention take advantage of the inventors' conception and observation that the C-terminus of the EGFR protein in which the tyrosine residues are phosphorylated can bind substrates (see below). By creating a labelled probe, wherein the tyrosines are phosphorylated with ^{32}P, the present inventors screened lambda gt11 cDNA expression libraries from various human tissues and identified numerous proteins which become associated with the phosphorylated C-terminus of the EGFR. The present inventors have termed the proteins discovered in this manner, "GRB" (for Growth factor Receptor Bound). The cloning methodology of the present invention has been designated, "CORT" (for Cloning Of Receptor Targets).

The method of the present invention is proposed as a novel approach having both generality and rapidity for the identification and cloning of target molecules for tyrosine kinases.

The present invention is thus directed to a method for detecting expression of a protein, which is a target of a receptor tyrosine kinase, by a cell harboring an expression vector, the protein being capable of binding to a tyrosine-phosphorylated polypeptide portion of the receptor tyrosine kinase, the method comprising:

(a) contacting the cell, an extract thereof, a lysate thereof, or a supernatant thereof with a solid phase carrier, causing the binding of the protein to the carrier;

(b) incubating the carrier-bound protein with the tyrosine-phosphorylated polypeptide, allowing the polypeptide to bind to the carrier-bound protein;

(c) removing materials not bound to the carrier; and

(d) detecting the presence or measuring the amount of the tyrosine-phosphorylated polypeptide bound to the carrier,

thereby detecting the expression of the protein.

In a preferred embodiment, the receptor is the epidermal growth factor receptor, the platelet-derived growth factor receptor, or the fibroblast growth factor receptor.

This method is preferably performed using a prokaryotic cell, most preferably a bacterial cell such as E. coli. The cell may also be eukaryotic, such as a yeast or a mammalian cell.

Preferably, the phosphorylated polypeptide is detectably labeled, such as with the radiolabel, ^{32}P.

The solid phase carrier is preferably a nitrocellulose membrane, to which are transferred proteins released from lysed bacterial cells when a lambda gt11 library is being screened.

The present invention also provides a method for mapping to a human chromosome a gene encoding a protein which is capable of binding to a tyrosine-phosphorylated polypeptide portion of a receptor tyrosine kinase molecule, the method comprising:

(a) infecting bacterial cells with a human gene expression library;

(b) detecting a clone expressing the protein using a method according to claim 1;

(c) sequencing the DNA of the clone; and

(d) mapping the sequence to a human chromosome.

The present invention is also directed to a polypeptide probe useful in the detection of the expression of a protein capable of binding to a tyrosine-phosphorylated polypeptide portion of a receptor tyrosine kinase. The probe comprises an amino acid sequence derived from the tyrosine-phosphorylated portion of the receptor molecule, or a functional derivative thereof, lacks the tyrosine kinase domain, and the sequence must contain at least one phosphotyrosine residue, preferably 4 or 5 phosphotyrosines. The probe should be detectably labeled, preferably with ^{32}P.

A preferred probe has between about 25 and 250 amino acid residues.

The probe of the present invention is useful for detecting target proteins for receptor tyrosine kinases including epidermal growth factor receptor, platelet-derived growth factor receptor, and fibroblast growth factor receptor.

The present invention also includes a method for preparing the above probe, comprising

(a) providing the receptor, or a genetically engineered receptor-like derivative in substantially pure form, wherein the receptor or the receptor-like derivative has both a tyrosine kinase domain and a tyrosine-phosphorylated domain, the tyrosine-phosphorylated domain including at least one tyrosine residue capable of being phosphorylated by the tyrosine kinase;

(b) incubating the receptor or receptor-like derivative with [gamma-^{32}P]adenosine triphosphate under conditions permitting phosphorylation of the tyrosine residue, causing phosphorylation of the tyrosine residue

thereby producing the probe. In a preferred embodiment, the method includes the step of:

(c) additionally treating the phosphorylated receptor molecule with an agent capable of cleaving the molecule between the tyrosine kinase domain and the tyrosine-phosphorylated domain.

A preferred cleaving agent is cyanogen bromide.

In another embodiment, the above method involves a genetically engineered receptor-like derivative which is a polypeptide encoded by a DNA molecule comprising a DNA sequence encoding tyrosine kinase, linked to a DNA sequence encoding a selective enzymatic cleavage site, linked to a DNA sequence encoding the tyrosine-phosphorylated domain, and wherein the agent is an enzyme capable of cleaving at this cleavage site. Preferred enzymes are Factor Xa and thrombin.

Also provided is a method for purifying from a complex mixture a protein which is capable of binding to a tyrosine-phosphorylated polypeptide portion of a receptor tyrosine kinase molecule, the method comprising:

(a) contacting the complex mixture with a solid phase carrier to which a probe is bound, allowing the protein to bind to the probe;

(b) removing materials not bound to the carrier; and

(c) eluting the bound protein from the carrier,

thereby purifying the protein.

The present invention is also directed to a protein, GRB-1, having the amino acid sequence shown in Figure 4. The invention includes a protein, GRB-2, which includes the amino acid sequence shown in Figure 17.

The invention is also directed to a DNA molecule encoding the GRB-1 protein, and a DNA molecule encoding the

GRB-2 protein. Included are DNA molecules encoding functional derivatives of these proteins. When the DNA molecule naturally occurs, it is substantially free of the nucleotide sequences with which it is natively associated. The DNA molecules of this invention may be expression vehicles, such as plasmids.

Also provided is a host transformed with each of the above DNA molecules.

The present invention also includes a process for preparing the GRB-1 or the GRB-2 protein substantially free of other proteins with which each is natively associated, or a functional derivative thereof, comprising:

(a) culturing a host cell capable of expressing the protein under culturing conditions,

(b) expressing the protein or functional derivative; and

(c) recovering the protein or functional derivative from the culture.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a filter blot pattern showing that the carboxy-terminus of the EGFR interacts with GAP-SH2 immobilized on nitrocellulose filters. Bacterially-expressed trpE/GAP-SH2 fusion protein or trpE as a control was spotted at various concentrations onto nitrocellulose filters. The filters were hybridized overnight with [^{32}P]-labelled C-terminal domain of the EGFR. Autoradiography was for 2 hours.

Figure 2 is a schematic diagram depicting the method of cloning of receptor targets (CORT). C-terminal domain of the EGFR is phosphorylated with radiolabelled phosphorous. Lambda gt11 library was plated at a density of 4 x 10^4 plaques per 150 ml plate. The plaques were overlaid with IPTG-impregnated nitrocellulose filters for 12 hours, after which the plaques were transferred to nitrocellulose and incubated with the labelled probe. Positive colonies are then selected for further analysis.

Figure 3 shows autoradiograms of phage expressing GRB-1 protein. A) Primary screen demonstrating one positive

signal (arrow) out of 40,000 phage plated. B) Plaque purification of phage expressing GRB-1. All plaques bound to the [^{32}P]-labelled C-terminal domain of the EGFR.

Figure 4 shows the DNA sequence and predicted amino acid sequence of GRB-1. The protein has 724 amino acid residues.

Figure 5 compares the sequences of the SH2 domains of GRB-1 with other proteins with similar motifs. A) SH2 domains of GRB-1, c-src, v-abl, bovine PLC-gamma, GAP, and V-crk. N and C refer to N-Terminal and C-terminal SH2 domains respectively. Conservation amino acid substitutions are as defined by Schwartz and Dayhoff: (A,G,P,S,T); (L,I,V,M); (D,E,N,Q); (K,R,H); (F,Y,W); and C. Bold letters identify those position were the same or a conservative amino acid substitution is present at 5 or more position. Boxes identify conserved motifs. B) A similar comparison of the SH3 domain of GRB-1.

Figure 6 is a schematic diagram comparing the structural organization of the SH2 and SH3 domains. The scheme includes known proteins containing SH2 and SH3 domains, such as c-src, v-crk, PLC-gamma, GAP1 and GRB-1.

Figure 7 is a Northern blot of monkey mRNA with GRB-1 probe. 5ug of poly (A)+ mRNA, obtained from various monkey tissue, was electrophoresed on 1.2%/2.2M agarose-formaldehyde gel. The blot was hybridized with a [32P]-nick translated DNA probe corresponding to the insert from clone ki4.

Figure 8 is a gel pattern showing that antibodies to GRB-1 immunoprecipitate a protein of 85 kDa from biosynthetically labelled cells. Cells were metabolically labelled with [^{35}S]methionine, after which lysates were prepared and immunoprecipitated with either immune (I) or preimmune (P) serum. The immunoprecipitated protein was separated on a 8% SDS/PAGE. Autoradiography was performed overnight. Cell lines used include human glioblastoma cell line, U1242, rat bladder carcinoma cell line, NBT-II and NIH-3T3 cells.

Figure 9 depicts several wild-type and mutant proteins used in the studies. (A) EGF receptor constructs with

their known or predicted autophosphorylation sites. Wild-type (W.T.), Kinase negative (K721A), and carboxy-terminal deletion (CD126), were immunoprecipitated from previously described transfected NIH373 cells expressing -300,000 EGF receptors. EGFR-C represents a deletion mutant containing the cytoplasmic domain of the EGF receptor produced by baculovirus-infected SF9 cells. (B) Structure of PLC-gamma and trpE/GAP SH2 proteins indicating location of the SH2 and SH3 domains and PLC-gamma tyrosine phosphorylation sites.

Figure 10 is a gel pattern showing association of PLC-gamma with EGFR mutants. Wild-type (HER14), carboxy-terminal deletion (DC126), or kinase-negative (K721A) EGFR were immunoprecipitated with anti-EGFR mAb108. Receptors were autophosphorylated with [gamma-^{32}P-ATP. Concomitantly EGFR-C was added to protein A-Sepharose beads alone or to immunoprecipitated K721A receptors either with or without ATP. After further washes to remove ATP, lysate from - 15 x 10 6 PLC-gamma overexpressing 3T-P1 cells was added and mixed for 90 min at 4 C. After washing to remove unbound PLC-gamma, proteins were separated on a 6% SDS-gel and transferred to nitrocellulose for immunoblotting. One eighth of the sample was utilized for anti-PTyr blotting, the remainder for anti-PLC-gamma blotting (exposure time 14 h).

Figure 11 is a gel pattern showing that phosphorylation of PLC-gamma reduces its binding to the EGF receptor. Full length EGFR was immunoprecipitated with mAb108, and allowed to autophosphorylate. Lysate from PLC-gamma overexpressing 3T-P1 cells was added and mixed for 90 min at 4°C. After binding, ATP was added to one half of the samples allowing the PLC-gamma molecules to be phosphorylated by the EGF receptor. SDS-PAGE sample buffer was then added to one half of the EGFR-PLC-gamma complexes (NO WASH, left panel) and directly loaded onto the 6% gel. The other half was washed three times with HNTG and then loaded on the gel (WASH, right panel. After running duplicate samples on SDS-PAGE, the proteins were transferred to nitrocellulose and probed with anti-PLC-gamma and [^{125}I]protein A. The bands were subsequently cut from the nitrocellulose and quantitated in a

gamma counter. After three washes with HNTG, 50±5% (Mean±SEM, n = 4) of the non-phosphorylated PLC-gamma remained bound to the EGFR while only 22±4% of the phosphorylated PLC-gamma remained (exposure time: 12h).

Figure 12 is a gel pattern showing binding of EGFR-C to trpE proteins. (A) EGFR-C (0.5 μg) was immunoprecipitated with antibody C and washed. $MnCl_2$ alone or $MnCl_2$ and ATP were then added to facilitate autophosphorylation. trpE or trpE/GAP SH2 (approximately 2 μg). The immunoprecipitates were separated on a 10% SDS-gel, transferred to nitrocellulose and immunoblotting was performed with anti-trpE. For comparison, about 0.1 μg of trpE or trpE/GAP SH2 lysate was loaded directly on to the gel (right panel of A). (B) trpE or trpE/GAP SH2 was immunoprecipitated with anti-trpE antibodies and washed. Phosphorylated or non-phosphorylated EGFR-C (0.5 μg) was then added and allowed to bind as above. After washing, samples were separated on a 10% gel, transferred to nitrocellulose and probed with antibody C. The two samples on the right represent 0.5 μg of phosphorylated and non-phosphorylated kinase loaded directly onto the gel (exposure time: 2 h).

Figure 13 is a gel pattern showing binding of trpE/GAP SH2 to wild-type and mutant EGFR. (A) Wild-type receptor (HER14) or the carboxy-terminal deletion CD126 receptor were immunoprecipitated with mAb 108. $MnCl_2$ alone or $MnCl_2$ and ATP were then added to the autophosphorylated half of the receptor-containing samples. One set of CD126 was also cross-phosphorylated with 0.5 μg of EGFR-C. trpE/GAP SH2 was then added for 90 min at 4°C and, after three more washes, loaded onto SDS-PAGE. After transfer to nitrocellulose, blots were probed with anti-trpE (left panel), anti-EGFR RK2 (center panel), or anti-PTyr (right panel). RK2 and anti-PTyr are both 1/8 of the total sample and were separated on 7% SDS-PAGE. The remaining sample was loaded on a 10% gel for the anti-trpE blot (exposure time 14 h). (B) Lysates from NIH3T3 2.2 cells containing no EGFR (3T3) or from cells with kinase-negative receptors (K21A) were immunoprecipitated with mAb108. To all immunoprecipitates, 0.5 μg of EGFR-C was added and then

$MnCl_2$ alone or $MnCl_2$ and ATP. trpE/GAP SH2 was added and samples prepared and immunoblotted as in (A) (exposure time: 19 h).

Figure 14 is a gel pattern showing binding of PLC-gamma and trpE/GAP SH2 to the CNBr cleaved C-terminal fragment of EGFR. EGFR-C (10 μg) was incubated in a Centricon 30 in 20 mM HEPES, pH 7.5 with 100 μg BSA as a carrier protein. The phosphorylated and non-phosphorylated EGFR-C were then each divided in two with one half being stored in buffer while the other half was cleaved with CNBr. The four samples either with or without ATP, and with or without CNBr were then each brought up in 500 ul 1% Triton X-100 lysis buffer, split in two, and immunoprecipitated with anti-C antibody. After washing the immunoprecipitates, lysates containing PLC-gamma or trpE/GAP SH2 were added. Immunoblotting was then performed on the samples as above with anti-trpE or anti-PLC-gamma. For the right panel, a fraction of the cleaved and uncleaved EGFR-C (0.1 μg) was loaded directly on the gel without immunoprecipitation and immunoblotted with RK2 (exposure time 14 h). The dark band seen in all lines of the anti-trpE blot runs at about 40 kDa (also seen in Figure 13) and represents [^{125}I]protein A binding to the heavy chain of the immunoprecipitating antibody.

Figure 15 is a gel pattern showing binding of the tyrosine phosphorylated C-terminal EGFR fragment to trpE/GAP SH2 but not to trpE. EGFR-C (5 μg) was autophosphorylated by the addition of [gamma-32P]ATPl. The phosphorylated EGFR-C was concentrated in a Centricon 30, and then cleaved with CNBr in 70% formic acid. One half of the sample (350,000 c.p.m.) was allowed to bind to trpE or trpE/GAP SH2 as in Figure 12B, washed and run on a 10% SDS-gel. (A) Binding of phosphorylated CNBr cleaved EGFR-C to trpE (B) Binding of phosphorylated CNBr cleaved EGFR-C to trpE GAP SH2 (C) 3000 c.p.m. of CNBr- cleaved EGFR-C (D) for comparison 3000 c.p.m. of cleaved EGFR-C (exposure time 20 h). EGFR 984/1186 indicates the sequence of the tyrosine autophosphorylated fragment generated by CNBr.

Figure 16 shows the partial nucleotide sequence and predicted amino acid sequence of GRB-2.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present inventors have developed a novel probe and method using this probe for rapid expression cloning of DNA encoding proteins which have the characteristic of binding to the tyrosine-phosphorylated C-terminal tail a receptor tyrosine kinase molecule, in particular the epidermal growth factor receptor (EGFR).

By the term "expression" is intended the transcription and translation of a DNA sequence encoding a protein, to yield a protein molecule. Expression cloning is a method wherein the DNA being cloned encodes a protein. The desired DNA, typically in the form of a cDNA library, is detected by means of its expression and direct detection of the protein which it encodes. Expression cloning systems are well-known in the art (see: Sambrook, J. et al. (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), which is hereby incorporated by reference). Typically, the protein is expressed from a library such as a lambda gt11 library, via infection of bacteria with the lambda gt11 expression vector. The expressed protein, released into the area of the bacteriophage plaque, is transferred to a nitrocellulose filter and typically detected using an antibody to stain the filter. For novel proteins, or proteins which are not available in sufficient quantities to prepare an antibody, or proteins which do not have a known ligand for affinity-based detection, such conventional expression cloning methods cannot be used.

The expression cloning method of the present invention may be easily applied to many other receptor systems. For example, certain target molecules bind to the tyrosine phosphorylated portion of PDGFR and the colony stimulating factor-1 (CSF-1) (Coughlin, S.R. et al., Science 243:1191-1194 (1989); Kazlauskas, A. et al., Cell 58:1121-1133 (1989); Shurtleff, S.A. et al., EMBO J. 9:2415-2421 (1990);

Reedjik, M. et al., Mol. Cell. Biol. 10:5601-5608 (1990)). In these two receptors, the tyrosine phosphorylation occurs in a kinase insert domain, rather than in the C-terminal domain as is the case with the EGFR. Therefore, specific probes utilizing the kinase insert domain, or a functional derivative thereof (defined below), of either the PDGFR or CSF-1 receptor can be similarly used for expression cloning. Similar probes can also be constructed for the fibroblast growth factor (FGF) receptor (which is tyrosine phosphorylated in the C-terminal domain) or the HER 2/neu receptor, both of the which are also able to interact with SH2 containing proteins such as PLC-gamma. In other receptors, such as the insulin receptor, tyrosine phosphorylation occurs in the kinase domain itself.

Thus, while it will be appreciated that different sites are tyrosine-phosphorylated in different proteins, e.g., the C-terminal domain in the EGFR, the kinase domain in insulin receptor, and a kinase domain insert in PDGFR, the present invention recognizes the common features of all these structures, the presence of one or more phosphotyrosine residues, and the ability of certain cellular proteins to bind on the basis of affinity to a polypeptide containing one or more phosphotyrosines. While reference will generally be made below to a probe which is a C-terminal domain, with reference to the EGFR, this language is not intended to be limiting and is intended to include all of the other alternative tyrosine-phosphorylated domains discussed above.

The methods and approach of the present invention can be applied to the cloning and identification of all target molecules which are capable of interacting in a specific manner with tyrosine phosphorylated polypeptides, such as the activated phosphorylated receptors described above. Additional proteins which bind to tyrosine-phosphorylated sequences, such as the tyrosine-specific phosphatases, e.g. R-PTPases (Sap, J. et al., Proc. Natl. Acad. Sci. USA 87:6112-6116 (1990); Kaplan, R. et al., Proc. Natl. Acad. Sci. USA 87:7000-7004 (1990)). The methods are also applicable in the cloning and identification of proteins which bind to

phosphorylated serine/threonine residues, as with serine/threonine-specific phosphatases.

Use of the probe of the present invention allows the rapid cloning of DNA and identification of the encoded proteins from a gene expression library. The method is particularly useful with a bacteriophage lambda gt11 library. Screening a human fetal brain lambda gt11 expression library has permitted the present inventors to clone two genes and to characterize the proteins they encode. One, termed GRB-1, was fully sequenced and found to encode novel human protein with a molecular weight of about 85 kDa which contained two SH2 domains and one SH3 domain (Figure 4 and Figure 5). GRB-2, which was partially sequenced, also contains unique SH2 and SH3 domains.

SH2 domains, such as in the GAP and PLC-gamma proteins, are responsible for the association of these proteins with the phosphorylated C-terminus of the EGFR (see Example VI, below). Thus, one function of SH2 domains is to juxtapose the intracellular portion of receptor tyrosine kinase molecules with their substrates to facilitate efficient tyrosine phosphorylation.

Detailed analysis of one of the cDNA clones of the present invention, GRB-1, identified using the methods of the invention, reveals a novel sequence containing two SH2 domains and one SH3 domain. This protein is expressed in various tissues and cell lines. Its predicted molecular weight, 85 kDa, is consistent with its migration upon sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

By the term "receptor tyrosine kinase" is intended a transmembrane protein having an extracellular receptor domain, and one or more intracellular domains, including a domain having tyrosine kinase enzymatic activity. Additional intracellular domains may have sequence homology to SH2. These molecules are well known in the art (Williams, L.T. et al., Science 243:1564-1570 (1989); Ullrich, A. et al., Cell 61:203-212 (1990); Carpenter, G. et al. J. Biol. Chem. 265: 7709-7712 (1990)).

The proteins which interact with, and which may be phosphorylated by, the receptor tyrosine kinases are referred to as "target" proteins for these receptors, in distinction from the "ligands" for these receptors, which bind to the extracellular receptor domain.

According to the present invention, the expression cloning method is performed directly on a gene expression library, such as lambda gt11. In a preferred embodiment, the DNA is human cDNA. More preferably, the DNA is human fetal brain DNA. Using such a source as the starting material for the cloning of human genes has a great advantage over the alternative means, in which a large amount of tissue is taken, and antibodies produced, or the protein purified and partially sequenced, and oligonucleotide probes prepared from this sequence and used to screen a genomic DNA or cDNA library. The advantage of bypassing these steps is of most relevance in the case of human genes, since tissue is generally not available in large quantities, with the exception of placenta.

The expression library is screened in a single step. Preferably, the lambda plaques are blotted onto a solid carrier, preferably nitrocellulose, allowing the transfer of library DNA-encoded proteins which are expressed in the infected bacteria and transferred to the carrier. This carrier is then incubated with the probe of the present invention, as described herein. The probe is allowed to bind to proteins which have the capability of binding to the tyrosine-phosphorylated polypeptide. Based on the label used in the probe, preferably the radioisotope ^{32}P, an appropriate detection system is used to identify the plaques containing the protein of interest. The phage in these plaques are then selected, and the DNA inserts in them can be re-cloned, excised and placed into other vectors, used for large scale expression of the protein, and the like, as is well-known in the art.

One of ordinary skill in the art will appreciate that the concentrations, times, temperatures can be varied depending on the precise nature of the system used, and will know how to vary the appropriate parameters without undue

experimentation. Furthermore, general methods in this area are set forth in Sambrook et al. (supra).

Materials of which solid phase carrier can be made include, but are not limited to, nitrocellulose, cellulose, paper, substituted polystyrenes, acrylonitriles, polycarbonate, polypentene, or silicone oxide.

The probe of the present invention is a tyrosine-phosphorylated polypeptide molecule derived from the C-terminal domain of a receptor tyrosine kinase, preferably the EGFR. The polypeptide can have between about 25 and about 250 amino acids in length. The probe can be a phosphorylated native sequence or a functional derivative thereof (defined below). Highly efficient phosphorylation is obtained by using the tyrosine kinase domain present on the receptor molecule to autophosphorylate the C-terminal region at between 1 and 5 tyrosine residues. Known methods and conditions (described in detail in Example I) are used to phosphorylate the tyrosine residues. A preferred substrate is [gamma-P^{32}-adenosine triphosphate). The source of the receptor molecule used as the source material to make the probe can include molecules chemically purified from tissues or cells, or molecules produced recombinant DNA methods.

When using recombinant techniques, a native receptor may be produced, or alternatively, a receptor-like derivative may be produced. A preferred receptor-like derivative includes the tyrosine kinase domain linked to the C-terminal domain. In another embodiment, the two domains may be produced as separate molecules, and mixed together to achieve tyrosine phosphorylation of the C-terminus-derived polypeptide.

The probe comprising a tyrosine-phosphorylated C-terminal portion of the receptor tyrosine kinase, as described herein can be produced by recombinant means in the form of a fusion protein.

As used herein, a "fusion protein" may refer to a fused protein comprising a bacterial protein and a polypeptide of interest such as a protein having an SH2 domain. Alternatively, a fusion protein may also be an artificially

constructed receptor tyrosine kinase-like derivative, wherein a DNA sequence encoding the tyrosine kinase domain has been linked to a selective enzymatic cleavage site, which, in turn, is linked to a receptor tyrosine kinase C-terminal domain having one or more tyrosine residues which can be phosphorylated by the kinase. Such a genetic construct encoding this type of "fusion protein" can be inserted into an expression vehicle and expressed in a bacterial or eukaryotic host. Once expressed, such a fusion protein can be allowed to autophosphorylate, wherein the kinase acts to phosphorylate the tyrosine residues in the C-terminal domain. Following this phosphorylation, use of the appropriate enzyme will cleave at the selective cleavage site, thus separating the N-terminal kinase from the C-terminal phosphorylated polypeptide, which can now serve as a probe.

Itakura et al. (Science 198:1056-1063 (1977)) and Riggs (U.S. Patent 4,366,246 (1982)) taught methods for bacterial expression of a foreign protein, in the form of a fusion protein, which prevents intracellular degradation of the foreign protein. Furthermore, overexpressed fusion proteins are frequently deposited as inclusion bodies in the bacterial cells, and are therefore easier to isolate and purify (Marston, Biochem. J. 240:1-12 (1986)). These references suggested alternate cleavage methods based on the concept of selective cleavage sites: Just as methionine could be inserted at the desired junction to serve as a target site for cyanogen bromide, so too could specific amino acid sites be introduced which could be attacked by proteolytic enzymes with amino acid specificity. Nagai et al. (Nature 309:810-812 (1984) disclosed the production of a desired gene product in a fusion protein by introducing a sequence of 4 amino acids which serve as a cleavage site for the blood coagulation factor Xa. Cleavage by Factor Xa also eliminated the extra N-terminal methionine encoded by the initiation codon which is present in most eukaryotic proteins expressed in E. coli by conventional methods. Other references describing the use of enzymatic cleavage sites recognized by Factor Xa or thrombin, for enhanced expression in bacteria and greater ease of

purification of the desired protein (Germino et al., Proc. Natl. Acad. Sci. USA 81:692-4696 (1984); Scholtissek et al., Gene 62:55-64 (1988); Smith et al., Gene 67:31-40 (1988); Knott et al., Eur. J. Biochem. 174:405-410 (1988); and Dykes et al., Eur. J. Biochem. 174:411-416 (1988)).

The term "selective cleavage site" refers to an amino acid residue or residues which can be selectively cleaved with either chemicals or enzymes and where cleavage can be achieved in a predictable manner. A selective enzymatic cleavage site is an amino acid or a peptide sequence which is recognized and hydrolyzed by a proteolytic enzyme. Examples of such sites include trypsin or chymotrypsin cleavage sites. In a preferred embodiment of this invention, the selective cleavage site is comprised of the sequence Ile-Glu-Gly-Arg, which is recognized and cleaved by blood coagulation factor Xa. In another embodiment, the selective cleavage site has the sequence Leu-Val-Pro-Arg, which is recognized and cleaved by thrombin.

In constructing the receptor-like derivative, an oligonucleotide sequence, 5' to the sequence coding for the enzyme recognition site can be included, and may vary in length. For example, in one embodiment, 13 nucleotides are situated between the codon for Ile (the start of the factor Xa recognition site) and the 3' end of the sequence encoding the tyrosine kinase domain.

Thus, in one embodiment of the present invention, the Ile-Glu-Gly-Arg sequence is introduced between the tyrosine kinase domain and the C-terminal domain. In another embodiment, the Leu-Val-Pro-Arg sequence is introduced. The proteins having this cleavage site are expressed in bacteria using standard methods. Thereafter, autophosphorylation of the C-terminal domain, preferably with [gamma32P] adenosine triphosphate, is allowed to occur, followed by selective cleavage of the tyrosine-phosphorylated C-terminal domain with the appropriate cleaving agent, e.g. factor Xa.

The present invention also provides a method for mapping a gene, preferably a human gene, which encodes a target protein for a receptor tyrosine kinase (such as a GRB

protein as defined herein), to a particular human chromosome. This method combines the new expression cloning method described herein with one of several known techniques for mapping a gene to a particular chromosome. Thus, according to the present invention, a clone, such as a lambda gt11 clone, containing a DNA insert encoding a GRB protein, is identified using the expression cloning methods disclosed herein. The insert may be further subcloned, if desired, using methods well-known in the art, and a probe constructed, either by direct labeling of the nucleic acid of the clone or by producing an oligonucleotide probe corresponding to a unique portion of the clone's sequence (see: Sambrook, J. et al. (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989)). This labeled probe can is then used in a hybridization assay with commercially available blots, such Chromosome Blots from Bios Corporation (New Haven, Connecticut) which contain DNA from a panel of human-hamster somatic cell hybrids (Kouri, R. E. et al., Cytogenet. Cell Genet. 51:1025 (1989)). By comparison of which human chromosomes remain in the human-hamster hybrid cell and the hybridization of the probe specific for the GRB gene of interest, the gene is mapped to a particular human chromosome. In this way, linkage is established to known human genes (or diseases caused by mutations therein) present on this chromosome. Using methods well-known in the art for finer mapping, e.g., using known human deletion mutations, the GRB gene can be mapped more precisely to other human genes.

The tyrosine-phosphorylated receptor tyrosine kinase C-terminal probe polypeptide of the present invention, as well as the GRB proteins of the present invention, and additional yet unknown GRB proteins which are discovered using the methods of this invention, are useful in methods for screening drugs and other agents which are capable of modulating cell growth control that occurs via signal transduction through receptor tyrosine kinases. By attaching a tyrosine-phosphorylated probe polypeptide or a GRB protein, or fragments thereof, to a solid phase carrier matrix, an affinity probe is created which can be used to isolate and

purify molecules from complex mixtures which are capable of binding to the affinity probe. Furthermore, such an affinity probe is useful for detecting the presence in a biological fluid of a molecule capable of binding the tyrosine-phosphorylated probe or the GRB protein. Similarly, chemical agents can be tested for their capacity to interact with the probe or GRB.

Methods for coupling proteins and peptides to the solid phase, the solid phase substances useful in these methods, and means for elution, are well known to those of skill in the art.

In the case of growth factor receptors which are receptor tyrosine kinases, including EDGFR, PDGFR and FGFR, tyrosine phosphorylation is linked to cell growth and to oncogenic transformation. Disruption of the action of a GRB in the cell may prevent or inhibit growth, and might serve as means to counteract development of a tumor. Furthermore, a mutation in the C-terminal portion of the receptor tyrosine kinase or the GRB, or a dysregulation in their mutual interactions, may promote susceptibility to cancer

The insulin receptor (InsR) is also a receptor tyrosine kinase, and tyrosine phosphorylation in cells bearing InsR is associated with normal physiological function. In contrast to the case of cell growth and cancer, disruption of normal interactions between of the tyrosine-phosphorylated portion of the receptor and the GRB would counteract insulin effects. Subnormal levels or activity of a GRB protein may act to remove a normal counterregulatory mechanisms. Perhaps overexpression or overactivity of a GRB protein could inhibit or totally prevent the action of insulin on cells, leading to diabetes (of an insulin-resistant variety). Thus susceptibility to diabetes may be associated with GRB-1 or GRB-2 protein dysregulation.

Therefore methods of the present invention for identifying normal or mutant GRB protein genes, or for detecting the presence or the amount of GRB-1 or GRB-2 in a cell can serve as methods for identifying susceptibility to cancer, diabetes, or other diseases associated with

alterations in cellular metabolism mediated by receptor tyrosine kinase pathways.

The present invention provides methods for evaluating the presence, and the level of normal or mutant GRB-1 or GRB-2 protein in a subject. Altered expression of these proteins, or presence of a mutant GRB protein, in an individual may serve as an important predictor of susceptibility to oncogenic transformation and the development of cancer. Alternatively, altered expression of GRB protein may serve as an important predictor of susceptibility to diabetes.

Oligonucleotide probes encoding various portions of the GRB protein are used to test cells from a subject for the presence DNA or RNA sequences encoding the GRB protein. A preferred probe would be one directed to the nucleic acid sequence encoding at least 4 amino acid residues, and preferably at least 5 amino acid residues of the GRB-1 or GRB-2 protein of the present invention. Qualitative or quantitative assays can be performed using such probes. For example, Northern analysis (see Example III, below) is used to measure expression of an GRB protein mRNA in a cell or tissue preparation.

Such methods can be used even with very small amounts of DNA obtained from an individual, following use of selective amplification techniques. Recombinant DNA methodologies capable of amplifying purified nucleic acid fragments have long been recognized. Typically, such methodologies involve the introduction of the nucleic acid fragment into a DNA or RNA vector, the clonal amplification of the vector, and the recovery of the amplified nucleic acid fragment. Examples of such methodologies are provided by Cohen et al. (U.S. Patent 4,237,224), Sambrook et al. (supra), etc.

Recently, an in vitro, enzymatic method has been described which is capable of increasing the concentration of such desired nucleic acid molecules. This method has been referred to as the "polymerase chain reaction or "PCR" (Mullis, K. et al., Cold Spring Harbor Symp. Quant. Biol.

51:263-273 (1986); Erlich H. et al., EP 50,424; EP 84,796, EP 258,017, EP 237,362; Mullis, K., EP 201,184; Mullis K. et al., US 4,683,202; Erlich, H., US 4,582,788; and Saiki, R. et al., US 4,683,194).

The polymerase chain reaction provides a method for selectively increasing the concentration of a particular nucleic acid sequence even when that sequence has not been previously purified and is present only in a single copy in a particular sample. The method can be used to amplify either single- or double-stranded DNA. The essence of the method involves the use of two oligonucleotide probes to serve as primers for the template-dependent, polymerase mediated replication of a desired nucleic acid molecule.

The precise nature of the two oligonucleotide probes of the PCR method is critical to the success of the method. As is well known, a molecule of DNA or RNA possesses directionality, which is conferred through the 5'-3' linkage of the phosphate groups of the molecule. Sequences of DNA or RNA are linked together through the formation of a phosphodiester bond between the terminal 5' phosphate group of one sequence and the terminal 3' hydroxyl group of a second sequence. Polymerase dependent amplification of a nucleic acid molecule proceeds by the addition of a 5' nucleotide triphosphate to the 3' hydroxyl end of a nucleic acid molecule. Thus, the action of a polymerase extends the 3' end of a nucleic acid molecule. These inherent properties are exploited in the selection of the oligonucleotide probes of the PCR. The oligonucleotide sequences of the probes of the PCR method are selected such that they contain sequences identical to, or complementary to, sequences which flank the particular nucleic acid sequence whose amplification is desired.

More specifically, the oligonucleotide sequences of the "first" probe is selected such that it is capable of hybridizing to an oligonucleotide sequence located 3' to the desired sequence, whereas the oligonucleotide sequence of the "second" probe is selected such that it contains an oligonucleotide sequence identical to one present 5' to the

desired region. Both probes possess 3' hydroxy groups, and therefore can serve as primers for nucleic acid synthesis.

In the PCR, the reaction conditions are cycled between those conducive to hybridization and nucleic acid polymerization, and those which result in the denaturation of duplex molecules. In the first step of the reaction, the nucleic acids of the sample are transiently heated, and then cooled, in order to denature any double-stranded molecules which may be present. The "first" and "second" probes are then added to the sample at a concentration which greatly exceeds that of the desired nucleic acid molecule. When the sample is incubated under conditions conducive to hybridization and polymerization, the "first" probe will hybridize to the nucleic acid molecule of the sample at a position 3' to the sequence to be amplified. If the nucleic acid molecule of the sample was initially double-stranded, the "second" probe will hybridize to the complementary strand of the nucleic acid molecule at a position 3' to the sequence which is the complement of the sequence whose amplification is desired. Upon addition of a polymerase, the 3' ends of the "first" and (if the nucleic acid molecule was double-stranded) "second" probes will be extended. The extension of the "first" probe will result in the synthesis of an oligonucleotide having the exact sequence of the desired nucleic acid. Extension of the "second" probe will result in the synthesis of an oligonucleotide having the exact sequence of the complement of the desired nucleic acid.

The PCR reaction is capable of exponential amplification of specific nucleic acid sequences because the extension product of the "first" probe, of necessity, contains a sequence which is complementary to a sequence of the "second" probe, and thus can serve as a template for the production of an extension product of the "second" probe. Similarly, the extension product of the "second" probe, of necessity, contains a sequence which is complementary to a sequence of the "first" probe, and thus can serve as a template for the production of an extension product of the "first" probe. Thus, by permitting cycles of polymerization,

and denaturation, a geometric increase in the concentration of the desired nucleic acid molecule can be achieved. Reviews of the PCR are provided by Mullis, K.B. (Cold Spring Harbor Symp. Quant. Biol. 51:263-273 (1986)); Saiki, R.K., et al. (Bio/Technology 3:1008-1012 (1985)); and Mullis, K.B., et al. (Meth. Enzymol. 155:335-350 (1987)).

In one embodiment, the invention is directed to a naturally occurring GRB-1 and GRB-2 proteins. In another embodiment, the invention is directed to recombinant GRB-1 and GRB-2 proteins. The invention provides the naturally occurring protein molecule substantially free of other proteins with which it is natively associated. "Substantially free of other proteins or glycoproteins" indicates that the protein has been purified away from at least 90 per cent (on a weight basis), and from even at least 99 per cent if desired, of other proteins and glycoproteins with which it is natively associated, and is therefore substantially free of them. That can be achieved by subjecting the cells, tissue or fluids containing the GRB-1 or GRB-2 protein to standard protein purification techniques such as immunoadsorbent columns bearing monoclonal antibodies reactive against the protein.

The nucleotide sequence of the GRB-1 gene, and the amino acid sequence of the GRB-1 protein, are shown in Figure 4. The partial nucleotide sequence of GRB-2 and the partial amino acid sequence, are shown in Figure 16.

In a preferred embodiment, GRB-1 or GRB-2, or an unknown GRB protein, can be isolated and purified using as an affinity probe, the probe of the present invention which is a tyrosine-phosphorylated C-terminal domain of a receptor tyrosine kinase, or a functional derivative thereof.

Alternatively, the purification can be achieved by a combination of standard methods, such as ammonium sulfate precipitation, molecular sieve chromatography, and ion exchange chromatography.

It will be understood that the GRB-1 and GRB-2 proteins of the present invention can be biochemically purified from a variety of cell or tissue sources. For

preparation of naturally occurring GRB protein, tissues such as mammalian placenta or brain are preferred.

Alternatively, because the gene for GRB1 and GRB-2 can be isolated or synthesized, the polypeptide can be synthesized substantially free of other proteins or glycoproteins of mammalian origin in a prokaryotic organism or in a non-mammalian eukaryotic organism, if desired. As intended by the present invention, a recombinant GRB-1 or GRB-2 molecule produced in mammalian cells, such as transfected COS, NIH-3T3, or CHO cells, for example, is either a naturally occurring protein sequence or a functional derivative thereof. Where a naturally occurring protein or glycoprotein is produced by recombinant means, it is provided substantially free of the other proteins and glycoproteins with which it is natively associated.

Alternatively, methods are well known for the synthesis of polypeptides of desired sequence on solid phase supports and their subsequent separation from the support. In particular, the tyrosine-phosphorylated C-terminal domain probe of the present invention, or a functional derivative thereof, can be synthesized using a peptide synthesis method wherein phosphotyrosine is provided in place of tyrosine, resulting in direct synthesis of the phosphorylated form of the polypeptide.

The present invention provides "functional derivatives" of the tyrosine-phosphorylated C-terminal domain polypeptide and or the GRB-1 and GRB-2 proteins.

By "functional derivative" is meant a "fragment," "variant," "analog," or "chemical derivative" of the GRB protein, which terms are defined below. A functional derivative retains at least a portion of the function of the native protein which permits its utility in accordance with the present invention.

A "fragment" of any of the proteins or polypeptides of the present invention refers to any subset of the molecule, that is, a shorter peptide.

A "variant" of the protein refers to a molecule substantially similar to either the entire peptide or a fragment

thereof. Variant peptides may be conveniently prepared by direct chemical synthesis of the variant peptide, using methods well- known in the art.

Alternatively, amino acid sequence variants of the peptide can be prepared by mutations in the DNA which encodes the synthesized peptide. Such variants include, for example, deletions from, or insertions or substitutions of, residues within the amino acid sequence. Any combination of deletion, insertion, and substitution may also be made to arrive at the final construct, provided that the final construct possesses the desired activity. Obviously, the mutations that will be made in the DNA encoding the variant peptide must not alter the reading frame and preferably will not create complementary regions that could produce secondary mRNA structure (see European Patent Publication No. EP 75,444).

At the genetic level, these variants ordinarily are prepared by site-directed mutagenesis (as exemplified by Adelman et al., DNA 2:183 (1983)) of nucleotides in the DNA encoding the peptide molecule, thereby producing DNA encoding the variant, and thereafter expressing the DNA in recombinant cell culture (see below). The variants typically exhibit the same qualitative biological activity as the nonvariant peptide.

An "analog" of the tyrosine-phosphorylated polypeptide or the GRB protein refers to a non-natural molecule substantially similar to either the entire molecule or a fragment thereof.

A "chemical derivative" of the tyrosine-phosphorylated polypeptide or the GRB protein contains additional chemical moieties not normally a part of the peptide. Covalent modifications of the peptide are included within the scope of this invention. Such modifications may be introduced into the molecule by reacting targeted amino acid residues of the peptide with an organic derivatizing agent that is capable of reacting with selected side chains or terminal residues.

Cysteinyl residues most commonly are reacted with alpha-haloacetates (and corresponding amines), such as chloroacetic acid or chloroacetamide, to give carboxymethyl or

carboxyamidomethyl derivatives. Cysteinyl residues also are derivatized by reaction with bromotrifluoroacetone, alpha-bromo- beta-(5-imidozoyl)propionic acid, chloroacetyl phosphate, N- alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloromercuribenzoate, 2-chloromercuri-4- nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

Histidyl residues are derivatized by reaction with diethylprocarbonate at pH 5.5-7.0 because this agent is relatively specific for the histidyl side chain. Para-bromophenacyl bromide also is useful; the reaction is preferably performed in 0.1 M sodium cacodylate at pH 6.0.

Lysinyl and amino terminal residues are reacted with succinic or other carboxylic acid anhydrides. Derivatization with these agents has the effect of reversing the charge of the lysinyl residues. Other suitable reagents for derivatizing alpha-amino-containing residues include imidoesters such as methyl picolinimidate; pyridoxal phosphate; pyridoxal; chloroborohydride; trinitrobenzenesulfonic acid; O-methylisourea; 2,4 pentanedione; and transaminase-catalyzed reaction with glyoxylate.

Arginyl residues are modified by reaction with one or several conventional reagents, among them phenylglyoxal, 2,3- butanedione, 1,2-cyclohexanedione, and ninhydrin. Derivatization of arginine residues requires that the reaction be performed in alkaline conditions because of the high pK_a of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine epsilon-amino group.

The specific modification of tyrosyl residues per se has been studied extensively, with particular interest in introducing spectral labels into tyrosyl residues by reaction with aromatic diazonium compounds or tetranitromethane. Most commonly, N-acetylimidizol and tetranitromethane are used to form O-acetyl tyrosyl species and 3-nitro derivatives, respectively.

Carboxyl side groups (aspartyl or glutamyl) are selectively modified by reaction with carbodiimides (R'-N-C-N-R') such as 1-cyclohexyl-3-(2-morpholinyl-(4-ethyl) carbodiimide or 1- ethyl-3-(4-azonia-4,4-dimethylpentyl) carbodiimide. Furthermore, aspartyl and glutamyl residues are converted to asparaginyl and glutaminyl residues by reaction with ammonium ions.

Glutaminyl and asparaginyl residues are frequently deamidated to the corresponding glutamyl and aspartyl residues. Alternatively, these residues are deamidated under mildly acidic conditions. Either form of these residues falls within the scope of this invention.

Derivatization with bifunctional agents is useful for cross-linking the peptide to a water-insoluble support matrix or to other macromolecular carriers. Commonly used cross-linking agents include, e.g., 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-hydroxysuccinimide esters, for example, esters with 4-azidosalicylic acid, homobifunctional imidoesters, including disuccinimidyl esters such as 3,3'-dithiobis(succinimidylpropionate), and bifunctional maleimides such as bis-N-maleimido-1,8-octane. Derivatizing agents such as methyl-3-[(p-azidophenyl)dithio]propioimidate yield photoactivatable intermediates that are capable of forming crosslinks in the presence of light. Alternatively, reactive water-insoluble matrices such as cyanogen bromide-activated carbo
hydrates and the reactive substrates described in U.S. Patent Nos. 3,969,287; 3,691,016; 4,195,128; 4,247,642; 4,229,537; and 4,330,440 are employed for protein immobilization.

Other modifications include hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the alpha-amino groups of lysine, arginine, and histidine side chains (T.E. Creighton, Proteins: Structure and Molecule Properties, W.H. Freeman & Co., San Francisco, pp. 79-86 (1983)), acetylation of the N-terminal amine, and, in some instances, amidation of the C-terminal carboxyl groups.

Such derivatized moieties may improve the solubility, absorption, biological half life, and the like. The moieties may alternatively eliminate or attenuate any undesirable side effect of the protein and the like. Moieties capable of mediating such effects are disclosed, for example, in Remington's Pharmaceutical Sciences, 16th ed., Mack Publishing Co., Easton, PA (1980)

The preferred bacterial host for this invention is E. coli. In other embodiments, other bacterial species can be used. In yet other embodiments, eukaryotic cells may be utilized, such as, for example, yeast, filamentous fungi, or the like. Use of these cell types are well known in the art. Any host may be used to express the protein which is compatible with replicon and control sequences in the expression plasmid. In general, vectors containing replicon and control sequences are derived from species compatible with a host cell are used in connection with the host. The vector ordinarily carries a replicon site, as well as specific genes which are capable of providing phenotypic selection in infected or in transformed cells. The expression of the fusion protein can also be placed under control with other regulatory sequences which may be homologous to the organism in its untransformed state.

This invention is also directed to an antibody specific for an epitope of the GRB-1 or the GRB-2 protein and the use of such an antibody to detect the presence of, or measure the quantity or concentration of, the GRB protein in a cell, a cell or tissue extract, or a biological fluid.

The term "antibody" is meant to include polyclonal antibodies, monoclonal antibodies (mAbs), chimeric antibodies, and anti-idiotypic (anti-Id) antibodies.

Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of animals immunized with an antigen.

Monoclonal antibodies are a substantially homogeneous population of antibodies to specific antigens. MAbs may be obtained by methods known to those skilled in the art. See, for example Kohler and Milstein, Nature 256:495-497

(1975) and U.S. Patent No. 4,376,110. Such antibodies may be of any immunoglobulin class including IgG, IgM, IgE, IgA, GILD and any subclass thereof. The hybridoma producing the mAbs of this invention may be cultivated in vitro or in vivo. Production of high titers of mAbs in vivo production makes this the presently preferred method of production. Briefly, cells from the individual hybridomas are injected intraperitoneally into pristane-primed BALB/c mice to produce ascites fluid containing high concentrations of the desired mAbs. MAbs of isotype IgM or IgG may be purified from such ascites fluids, or from culture supernatants, using column chromatography methods well known to those of skill in the art.

Chimeric antibodies are molecules different portions of which are derived from different animal species , such as those having variable region derived from a murine mAb and a human immunoglobulin constant region. Chimeric antibodies and methods for their production are known in the art (Cabilly et al, Proc. Natl. Acad. Sci. USA 81:3273-3277 (1984); Morrison et al., Proc. Natl. Acad. Sci. USA 81:6851-6855 (1984); Boulianne et al., Nature 312:643-646 (1984); Cabilly et al., European Patent Application 125023 (published November 14, 1984); Neuberger et al., Nature 314:268-270 (1985); Taniguchi et al., European Patent Application 171496 (published February 19, 1985); Morrison et al., European Patent Application 173494 (published March 5, 1986); Neuberger et al., PCT Application WO 86/01533, (published March 13, 1986); Kudo et al., European Patent Application 184187 (published June 11, 1986); Morrison et al., European Patent Application 173494 (published March 5, 1986); Sahagan et al., J. Immunol. 137:1066-1074 (1986); Robinson et al., International Patent Publication #PCT/US86/02269 (published 7 May 1987); Liu et al., Proc. Natl. Acad. Sci. USA 84:3439-3443 (1987); Sun et al., Proc. Natl. Acad. Sci. USA 84:214-218 (1987); Better et al., Science 240:1041- 1043 (1988)). These references are hereby incorporated by reference.

An anti-idiotypic (anti-Id) antibody is an antibody which recognizes unique determinants generally associated with

the antigen-binding site of an antibody. An Id antibody can be prepared by immunizing an animal of the same species and genetic type (e.g. mouse strain) as the source of the mAb with the mAb to which an anti-Id is being prepared. The immunized animal will recognize and respond to the idiotypic determinants of the immunizing antibody by producing an antibody to these idiotypic determinants (the anti-Id antibody).

The anti-Id antibody may also be used as an "immunogen" to induce an immune response in yet another animal, producing a so-called anti-anti-Id antibody. The anti-anti-Id may be epitopically identical to the original mAb which induced the anti-Id. Thus, by using antibodies to the idiotypic determinants of a mAb, it is possible to identify other clones expressing antibodies of identical specificity.

Accordingly, mAbs generated against the GRB protein of the present invention may be used to induce anti-Id antibodies in suitable animals, such as BALB/c mice. Spleen cells from such immunized mice are used to produce anti-Id hybridomas secreting anti-Id mAbs. Further, the anti-Id mAbs can be coupled to a carrier such as keyhole limpet hemocyanin (KLH) and used to immunize additional BALB/c mice. Sera from these mice will contain anti-anti-Id antibodies that have the binding properties of the original mAb specific for a GRB protein epitope.

The anti-Id mAbs thus have their own idiotypic epitopes, or "idiotopes" structurally similar to the epitope being evaluated, such as GRB protein-α.

The term "antibody" is also meant to include both intact molecules as well as fragments thereof, such as, for example, Fab and $F(ab')_2$, which are capable of binding antigen. Fab and $F(ab')_2$ fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody (Wahl et al., J. Nucl. Med. 24:316-325 (1983)).

It will be appreciated that Fab and $F(ab')_2$ and other fragments of the antibodies useful in the present invention may be used for the detection and quantitation of

GRB protein according to the methods disclosed herein for intact antibody molecules. . Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce $F(ab')_2$ fragments).

An antibody is said to be "capable of binding" a molecule if it is capable of specifically reacting with the molecule to thereby bind the molecule to the antibody. The term "epitope" is meant to refer to that portion of any molecule capable of being bound by an antibody which can also be recognized by that antibody. Epitopes or "antigenic determinants" usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and have specific three dimensional structural characteristics as well as specific charge characteristics.

An "antigen" is a molecule or a portion of a molecule capable of being bound by an antibody which is additionally capable of inducing an animal to produce antibody capable of binding to an epitope of that antigen. An antigen may have one, or more than one epitope. The specific reaction referred to above is meant to indicate that the antigen will react, in a highly selective manner, with its corresponding antibody and not with the multitude of other antibodies which may be evoked by other antigens.

The antibodies, or fragments of antibodies, useful in the present invention may be used to quantitatively or qualitatively detect the presence of cells which express the GRB protein. This can be accomplished by immunofluorescence techniques employing a fluorescently labeled antibody (see below) coupled with light microscopic, flow cytometric, or fluorimetric detection.

The antibodies (of fragments thereof) useful in the present invention may be employed histologically, as in immunofluorescence or immunoelectron microscopy, for <u>in situ</u> detection of GRB proteins. <u>In situ</u> detection may be accomplished by removing a histological specimen from a patient, and providing the a labeled antibody of the present invention to such a specimen. The antibody (or fragment) is

preferably provided by applying or by overlaying the labeled antibody (or fragment) to a biological sample. Through the use of such a procedure, it is possible to determine not only the presence of the GRB protein but also its distribution on the examined tissue. Using the present invention, those of ordinary skill will readily perceive that any of a wide variety of histological methods (such as staining procedures) can be modified in order to achieve such in situ detection.

Such assays for GRB protein typically comprises incubating a biological sample, such as a biological fluid, a tissue extract, freshly harvested cells such as lymphocytes or leucocytes, or cells which have been incubated in tissue culture, in the presence of a detectably labeled antibody capable of identifying GRB protein, and detecting the antibody by any of a number of techniques well-known in the art.

The biological sample may be treated with a solid phase support such as nitrocellulose, or other solid support which is capable of immobilizing cells, cell particles or soluble proteins. The support may then be washed with suitable buffers followed by treatment with the detectably labeled GRB protein-specific antibody. The solid phase support may then be washed with the buffer a second time to remove unbound antibody. The amount of bound label on said solid support may then be detected by conventional means.

By "solid phase support" is intended any support capable of binding antigen or antibodies. Well-known supports, or carriers, include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite. The nature of the carrier can be either soluble to some extent or insoluble for the purposes of the present invention. The support material may have virtually any possible structural configuration so long as the coupled molecule is capable of binding to an antigen or antibody. Thus, the support configuration may be spherical, as in a bead, or cylindrical, as in the inside surface of a test tube, or the external surface of a rod. Alternatively, the surface may be flat such as a sheet, test strip, etc. Preferred

supports include polystyrene beads. Those skilled in the art will know many other suitable carriers for binding antibody or antigen, or will be able to ascertain the same by use of routine experimentation.

The binding activity of a given lot of anti-GRB-1 and anti-GRB-2 antibody may be determined according to well known methods. Those skilled in the art will be able to determine operative and optimal assay conditions for each determination by employing routine experimentation.

Other such steps as washing, stirring, shaking, filtering and the like may be added to the assays as is customary or necessary for the particular situation.

One of the ways in which the GRB-specific antibody can be detectably labeled is by linking the same to an enzyme and use in an enzyme immunoassay (EIA). This enzyme, in turn, when later exposed to an appropriate substrate, will react with the substrate in such a manner as to produce a chemical moiety which can be detected, for example, by spectrophotometric, fluorimetric or by visual means. Enzymes which can be used to detectably label the antibody include, but are not limited to, malate dehydrogenase, staphylococcal nuclease, delta-5-steroid isomerase, yeast alcohol dehydrogenase, alpha-glycerophosphate dehydrogenase, triose phosphate isomerase, horseradish peroxidase, alkaline phosphatase, asparaginase, glucose oxidase, beta-galactosidase, ribonuclease, urease, catalase, glucose-6-phosphate dehydrogenase, glucoamylase and acetylcholinesterase. The detection can be accomplished by colorimetric methods which employ a chromogenic substrate for the enzyme. Detection may also be accomplished by visual comparison of the extent of enzymatic reaction of a substrate in comparison with similarly prepared standards.

Detection may be accomplished using any of a variety of other immunoassays. For example, by radioactively labeling the antibodies or antibody fragments, it is possible to detect R- PTPase through the use of a radioimmunoassay (RIA). A good description of RIA may be found in Laboratory Techniques and Bio chemistry in Molecular Biology, by Work, T.S., et al.,

North Holland Publishing Company, NY, (1978) with particular reference to the chapter entitled "An Introduction to Radioimmune Assay and Related Techniques" by Chard, T., incorporated by reference herein. The radioactive isotope can be detected by such means as the use of a gamma counter or a scintillation counter or by autoradiography.

It is also possible to label the antibody with a fluorescent compound. When the fluorescently labeled antibody is exposed to light of the proper wave length, its presence can then be detected due to fluorescence. Among the most commonly used fluorescent labelling compounds are fluorescein isothiocyanate, rhodamine, phycoerythrin, phycocyanin, allophycocyanin, o- phthaldehyde and fluorescamine.

The antibody can also be detectably labeled using fluorescence emitting metals such as ^{152}Eu, or others of the lanthanide series. These metals can be attached to the antibody using such metal chelating groups as diethylenetriaminepentaacetic acid (DTPA) or ethylenediaminetetraacetic acid (EDTA).

The antibody also can be detectably labeled by coupling it to a chemiluminescent compound. The presence of the chemiluminescent-tagged antibody is then determined by detecting the presence of luminescence that arises during the course of a chemical reaction. Examples of particularly useful chemiluminescent labeling compounds are luminol, isoluminol, theromatic acridinium ester, imidazole, acridinium salt and oxalate ester.

Likewise, a bioluminescent compound may be used to label the antibody of the present invention. Bioluminescence is a type of chemiluminescence found in biological systems in which a catalytic protein increases the efficiency of the chemiluminescent reaction. The presence of a bioluminescent protein is determined by detecting the presence of luminescence. Important bioluminescent compounds for purposes of labeling are luciferin, luciferase and aequorin.

The antibody molecules of the present invention may be adapted for utilization in an immunometric assay, also known as a "two-site" or "sandwich" assay. In a typical

immunometric assay, a quantity of unlabeled antibody (or fragment of antibody) is bound to a solid support and a quantity of detectably labeled soluble antibody is added to permit detection and/or quantitation of the ternary complex formed between solid-phase antibody, antigen, and labeled antibody.

Typical, and preferred, immunometric assays include "forward" assays in which the antibody bound to the solid phase is first contacted with the sample being tested to extract the antigen from the sample by formation of a binary solid phase antibody-antigen complex. After a suitable incubation period, the solid support is washed to remove the residue of the fluid sample, including unreacted antigen, if any, and then contacted with the solution containing an unknown quantity of labeled antibody (which functions as a "reporter molecule"). After a second incubation period to permit the labeled antibody to complex with the antigen bound to the solid support through the unlabeled antibody, the solid support is washed a second time to remove the unreacted labeled antibody.

In another type of "sandwich" assay, which may also be useful with the antigens of the present invention, the so-called "simultaneous" and "reverse" assays are used. A simultaneous assay involves a single incubation step as the antibody bound to the solid support and labeled antibody are both added to the sample being tested at the same time. After the incubation is completed, the solid support is washed to remove the residue of fluid sample and uncomplexed labeled antibody. The presence of labeled antibody associated with the solid support is then determined as it would be in a conventional "forward" sandwich assay.

In the "reverse" assay, stepwise addition first of a solution of labeled antibody to the fluid sample followed by the addition of unlabeled antibody bound to a solid support after a suitable incubation period is utilized. After a second incubation, the solid phase is washed in conventional fashion to free it of the residue of the sample being tested and the solution of unreacted labeled antibody. The

determination of labeled antibody associated with a solid support is then determined as in the "simultaneous" and "forward" assays.

EXAMPLE I

A study was performed to determine the detectability of binding of the C-terminal domain of EGFR to a protein containing the SH2 domain immobilized on nitrocellulose filters. For this purpose, the binding of the C-terminal domain to a bacterially expressed fusion protein was assessed (see Figure 1).

A. Isolation and Labelling of the Carboxyterminal Domain of the EGFR

The intracellular portion of the EGFR, which includes the tyrosine kinase domain and the carboxy terminal domain, was purified from recombinant baculovirus which expressed cDNA complementary to the intracellular domain of the human EGFR, as described previously (Hsu, C-Y. et al., Cell Growth and Differentiation 1:191-200 (1990)). The recombinant protein (2 μg) was then phosphorylated with [gamma-^{32}P]ATP (200 μCi, 6000 Ci/Mmol)., at 4°C in HNTG (20 mM HEPES, pH 7.5, 150mM NaCl, 0.1% Triton X-100, and 10% glycerol) buffer which contained 5mM $MnCl_2$. In order to remove unincorporated [y-32P] ATP, the phosphorylated kinase was diluted to 1 ml with 20 mM HEPES, pH 7.5, containing 100 μg BSA and then concentrated in a Centricon-10 to a volume cf 50 μl. This procedure was repeated 3 times resulting in the removal of >99% of the unincorporated ATP. To separate the C-terminal domain from the kinase domain, the concentrated protein was then digested with cyanogen bromide (CNBr) in 70% formic acid for 14 hours at room temperature (see also Example VI, below). Samples were then washed three times with water, dried and resuspended in binding buffer to a concentration of 2 x 10^6 cpm/ml.

B. Binding of the C-terminal Domain of the EGFR to Bacterially Expressed TrpE/GAP-SH2 Fusion Protein Immobilized on Nitrocellulose

TrpE and TrpE/GAP-SH2 were obtained from the laboratory of Dr. Tony Pawson and prepared as previously described (Moran, M.F. et al., Proc. Natl. Acad. Sci. USA 87:8622-8626 (1990)). Filter binding studies were performed according to published methods (Schneider, W.J. et al., Proc. Natl. Acad. Sci. 76:5577-5581 (1979); Daniel, T.O. et al., J. Biol. Chem. 258:4606-4611 (1983)) with minor modifications. Various concentrations of either bacterially expressed TrpE fusion protein or bacterial protein alone were spotted onto nitrocellulose filters. After blocking the filters for 1 hour at 4°C in PBS containing 5% Carnation dry milk, $_{32}$P-labelled C-terminal domain of the EGFR was added and incubation was continued overnight at 4°C. After 24 hours, the nitrocellulose filters were washed 3 times at room temperature with PBS containing 0.2% Triton X-100. The filters were dried and exposed to Kodak XAR-5 film at -80°C.

C. Results

The above method permitted detection of specific binding of the EGFR C-terminal domain to less than 5 ng of a bacterially expressed GAP-SH2 fusion protein. The binding was specific, since it required tyrosine phosphorylation of the probe and did not occur when irrelevant proteins were applied to nitrocellulose filters.

The demonstration that the EGFR C-terminal domain could bind specifically to an SH2-containing protein immobilized on nitrocellulose filters encouraged the present inventors to apply this approach to the screening of lambda gt11 expression libraries with the goal of identifying novel EGFR binding proteins.

EXAMPLE II

Screening of Expression Libraries and Isolation of a cDNA Clone Encoding a Novel SH2-Containing Protein

The tyrosine phosphorylated C-terminal tail of the EGFR was used as a probe to screen expression libraries from

several different human tissues as described above. The approach to screening is outlined in Figure 2. Numerous positive clones have been identified so far using this approach, of which two have been analyzed in detail.

A. Screening of cDNA Library

A lambda gt11, library, constructed from mRNA isolated from human brain stem, was obtained from M. Jaye. To screen the library, lambda gt11 phage were plated at a density sufficient to produce 4 X 10^4 plaques per 150 mm agar plate. A total of six plates were initially screened. After incubation of the plates for 4 hours at 42°C, the plates were overlaid with nitrocellulose filters which had been impregnated with isopropyl-B-D-thiogalactopyranoside (IPTG), as previously described (MacGregor, P.F. et al., Oncogene 5:451-458 (1990)). Incubation was continued overnight at 37°C. The filters were then removed, washed with tBST (10 mM Tris-HCl, pH8, 150 mM NaCl, and 0.05% triton X-100) at room temperature, and then blocked in HBB (20 mM HEPES, pH 7.5, 5 mM Mg/Cl, 1 mM KCl) buffer containing 5% carnation dry milk for 1 hour at 4°C, as described (MacGregor et al., supra). Following blocking, labelled tyrosine phosphorylated carboxy-terminus (C-terminus) probe was added at a concentration of 1.6 X 10-4 μg/ml, and incubation was continued overnight. The filters were then washed 3 times at room temperature in PBS containing 0.2% Triton X-100. Filters were dried and exposed to Kodak XAR-5 film at -80°C.

Agar plugs, corresponding to the positive clones, were collect from the plates and placed in 1 ml of SM media. After allowing the phages to diffuse from the agar, the phages were replated and rescreened as described above. Those phages that demonstrated enrichment on subsequent screening were isolated and sequence. Lambda gt11 phage DNA was isolated by the plate lysate method according to Maniatis et al., and subcloned into EcoRI-digested M13 MP19 (Maniatis et al., 1982). Single stranded DNA was isolated and sequenced by the dideoxy chain termination method using the Sequenase DNA sequencing kit (United States Biochemical).

In one experiment, 240,000 pfu from a human brainstem lambda gt11 library were screened. A single plaque, clone ki4 (Figure 3A) was isolated. On subsequent screening this clone demonstrated enrichment, and on tertiary screening all plaques bound the probe (Figure 3B). Clone ki4 contained an insert of about 900 nucleotides, which, upon induction of the lac promoter with IPTG, produced a fusion protein which could bind the EGFR. The size of the fusion protein predicted that the cDNA insert coded for a protein of about 300 amino acids, which was the size expected if the cDNA contained a single large open reading frame. To analyze clone ki4 in more detail, DNA was isolated and the EcoRI fragment, corresponding to the human cDNA insert, was subcloned into M13 and sequenced. Translation of the sequence from this insert demonstrated a single large open reading frame which, upon analysis using the Genbank database, was found to contain a single stretch of about 100 amino acids with sequence homology to SH2 domains of other known proteins (Figures 4 and 5A). However, in other regions, no sequence homology was noted. Thus, using this screening approach, a new SH2-containing protein which could bind to the EGFR was identified,.

B. Isolated of Full Length cDNA

The initial clone isolated encoded for an SH2 domain, but did not contain the 3' or 5' ends of the gene. To isolated the full length cDNA, the library was rescreened using DNA isolated from the initial positive phage. DNA, from recombinant M13 bacteriophage which expressed the positive clone, was amplified using a thermal cycler, Taq1 polymerase and oligonucleotides complementary to the EcoR1 flanking regions of the M13 vector. In order to obtain clones containing more 5' sequence in information, a second amplified DNA product, corresponding to the most 5' 250 nucleotides of the initial isolated phage, was also generated by using oligonucleotides complementary to sequences at both ends of this region. [^{32}P]-labelled DNA probes were then prepared by nick translation of the amplified products.

To rescreen the cDNA library, the library was replated as described above. After incubation of the plates for 8 hours at 37°C, the plates were cooled for 1 hour at 4°C following which the phage DNA was transferred to nitrocellulose filters. The filters were denatured in a solution of 0.2 N NaOH and 1.5 M NaCl and then baked in vacuo for 2 hours at 80°C (Sambrook, J. et al., (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989)). After prehybridization of the filters for 1 hour at 42°C, 32P-labelled DNA probe was added and hybridization was continued overnight at 42°C in a solution containing 5X Denhardt's, 50% formamide, 5X SSC, 0.1% SDS, 200 mM Tris-HCl, pH 7.6 and 100 μg/ml salmon sperm DNA. The filters were then washed in a solution containing 0.1X SSC and 0.1% SDS, dried and exposed to Kodak XAR-5 film at -70°C. Positive clones were then isolated and sequenced as described above.

Since the insert from clone ki4 lacked the 3' and 5' ends of the gene, the library was rescreened using two DNA probes which were generated by amplifying DNA from clone ki4. This approach enabled the identification of five additional clones. Three of the clones extended 3' from the initial clone ki4, two of which, clones, ki2.2 and ki2.4, contained a polyadenylation signal and a long 3' untranslated region (>1000 nucleotides). In addition, these clones encoded a protein which contained a second SH2 domain (Figures 4 and 5A).

The other two clones, ki3.0 and ki5.3, extended 5' from clone ki4. Both clones contained long open reading frames and an AUG codon which met the translation initiation criteria as defined by Kozak (Kozak, M. J. Cell. Biol. 108:229-241 (1989)). However, only clone ki3.0, when translated into protein and compared with known sequences in Genbank, was found to contain a domain of 50 amino acids which was homologous to SH3 domains present in other known proteins. The predicted molecular weight of the full length protein encoded by the overlapping clones, ki2.2 and ki3.0, was about 84 kDa. This new protein was termed GRB-1.

EXAMPLE III

GRB-1 Protein Contains SH2 and SH3 domains

Analysis of the GRB-1 protein sequence by comparison to sequences in the Genbank database revealed the presence of two stretches of about 100 amino acids, starting at amino acids 333 and 624, with sequence homology to SH2 domains of other proteins known to interact with the EGFR (Figure 5A). While GRB-1 displayed striking homology to other SH2 domains at the protein level, it revealed no significant homology at the DNA level. GRB-1 also contained a segment of about 50 amino acids, located in the N-terminal region, which had sequence homology to SH3 domains (Figure 4 and 5B).

A comparison of the structural organization of GRB-1 with several other SH2/SH3 containing proteins is shown in Figure 6. It is apparent from this scheme that the localization of the SH2 and SH3 domains vary from protein to protein. Despite this there are certain similarities and differences among these SH2 containing proteins. GRB-1 is similar to some other substrates which have been found to interact with the EGFR, such as PLC-gamma and GAP, in that GRB-1 contains two SH2 domains and a single SH3 domain. However, unlike these substrates, GRB-1 contains no homology to any known catalytic domain, and in this regard resembles the protein encoded by the avian sarcoma virus, v-crk.

Out side of these regions there was no sequence homology with other protein sequences present in Genbank. In particular, GRB-1 lacked a consensus ATP-binding domain, and did nod display sequence homology with any serine/threonine kinase or tyrosine kinase.

The SH2 domain is thought to provide a common motif by which enzymatically distinct signalling molecules can be coupled to activated receptors with tyrosine kinase activity (Moran, M.F. et al., Proc. Natl. Acad. Sci. USA 87:8622-8626 (1990); Anderson, D. et al., Science 250:979-982 (1990)).

The presence of SH2 domains in GRB-1 (Figure 4) and in GRB-2 further reinforces the importance of this domain in

mediating the interaction of these proteins with the C-terminal tail of the EGFR. Moreover, since many proteins capable of interacting with tyrosine receptors remain to be identified, this suggests that additional members of this protein family remain to be discovered.

In addition to containing two SH2 domains, GRB-1 also contains an SH3 domain. The SH3 domain is a non-catalytic domain of about 50 amino acid residues which is shared among many SH2-containing proteins. Since SH3 domains are also found in cytoskeletal proteins, such as spectrin and fodrin, the function of this domain could be to localize these proteins to the membrane or submembrane cytoskeleton where they would interact with other molecules.

Comparison of the deduced amino acid sequence of GRB-1 with the protein product encoded by the avian oncogene v-crk may shed light on GRB-1 function. The gene v-crk encodes a protein which is composed primarily of a viral gag protein fused to an SH2 and SH3 domain (Mayer, B.J. et al., Nature 332:272-275 (1988)). Both GRB-1 and the $p47^{gag-crk}$ protein have no homology with any known catalytic domains. However, chicken embryo fibroblasts transformed with $p47^{gag-crk}$ display elevated levels of phosphotyrosine-containing proteins (Mayer, B.J. et al., supra; Proc. Natl. Acad. Sci. USA 87:2638-2642 (1990); Matsuda, M. et al., Science 248:1537-1539 (1990)).

Since the v-crk product has been shown to bind several phosphotyrosine-containing proteins in v-crk transformed cells, it may be that the function of c-crk is to act as a bridge between kinases and substrates. In this regard, it is intriguing that GRB-1, like GAP and PLC-gamma, contains two SH2 domains, the combination of which may be ideally suited for linking other proteins to activated tyrosine kinase receptors.

EXAMPLE IV

Northern Analysis of GRB-1 Expression

A. Methods

Total cellular RNA was prepared from monkey tissue by the guanidinium isothiocyanate/cesium chloride method described by Sambrook, J. et al., (supra). Poly (A)+ RNA was prepared by oligo(dT) cellulose chromatography. For Northern analysis, RNA was size fractionated by electrophoresis in a 1.2% agarose/2.2M formaldehyde gel, transferred onto a nylon membrane by capillary action and baked at 80°C for 2 hours. Following prehybridization, the blot was hybridized with a [^{32}P]-nick-translated DNA probe which was prepared as descried above. Hybridization was carried out overnight at 42°C in the presence of 50% formamide, 5X SSC, 0.1% SDS, and 5X Denhardt's. The membrane was then washed in 0.1X SSC, 0.1% SDS at 42°C, and exposed to Kodak XAR film at -70°C for 12 hours using an intensifying screen.

B. Results

To test for the expression of mRNA corresponding to the newly isolated cDNA, Northern blot analysis of different monkey tissue mRNA, probed with DNA corresponding to the insert from clone ki4, demonstrated the presence of two major bands of 4.4 kb and 7.0 kb in most tissues examined (Figure 7). Expression was highest in the brain, with heart, spleen, liver and thymus displaying decreasing levels of expression. The 4.4 kb message corresponds to the expected size of the transcript which would encode the isolated clones. In contrast to the 4.4 and 7.0 kb transcripts observed in most tissues, the skin contained two slightly smaller sized mRNAs of 3.6 and 6.6 kb.

The 3.6, 6.6 and 7.0 kb transcripts may represent alternatively spliced forms of mRNA, or may encode for distinct but related mRNA species.

EXAMPLE V

Production of anti-GRB-1 Antibodies and Analysis of GRB-1 Fusion Protein

A. Methods

Polyclonal antibodies were produced by immunizing rabbits with the ß-galactosidase fusion protein expressed by the initial isolated phage clone, ki4. E. coli CAG 456 bacteria (obtained from Dr. Michael Snyder, Yale University) were infected with recombinant phage ki4 at a multiplicity-of-infection of 10 and ß-galactosidase fusion protein was recovered from the protein pellet after 1.5 hours. Protein extracts were prepared, separated on a 6% SDS-gel, and the band corresponding to the fusion protein excised from gel and used for immunization.

Human glioblastoma cell line U1242, rat bladder carcinoma cell line NBT II, and NIH3T3 cells were grown to confluence in DMEM medium supplemented with 10% fetal bovine serum. Cells were labelled with [^{35}S]-methionine (50 μCi/ml) in 0.5% fetal bovine serum and lysed after 12 hours as previously described (Margolis, B. et al., Cell 57:1101-1107 (1989)). After immunoprecipitation with 10 μl of antibody coupled to protein A-Sepharose, the beads were washed three times with a solution containing 20mM HEPES, pH 7.5, 300mM NaCl, 10% glycerol, 1% Triton X-100, 0.1% SDS, and 1% sodium deoxycholate. After boiling in sample buffer proteins were separated on a 8% SDS-gel.

B. Results

Polyclonal antibodies were raised against the ß-galactosidase fusion protein expressed by the initial isolated phage. Immunoprecipitation experiments, using biosynthetically labelled cells, demonstrated that these antibodies recognized an 85 kDa protein in three different cell lines (Figure 8, lanes designated "I"). Recognition of the 85 kDa protein by this antiserum was specific since preimmune serum did not recognize this protein (lanes designated "P"). These results provided support for the predicted molecular weight based on the amino acid sequence of cloned GRB-1.

C. Discussion

The finding that the gene for GRB-1 encodes for a protein with an expected molecular weight of 85 kDa, together with the demonstration that antibodies to GRB-1 immunoprecipitated an 85 kDa protein from three different cell lines, suggest that GRB-1 may represent a particular protein which had previously been shown to associate with activated growth factor receptors, namely p85. While the exact function of p85 was unknown, it was presumed to be phosphatidylinositol (PI3)-kinase, since PI3-kinase activity copurified with an 85 kDa protein found in PDGF- stimulated as well as middle T-antigen (MTAg)-transformed cells (Kaplan, D.R. Cell 50:1021-1029 (1987); Whitman, M. et al., Nature 315:239-242 (1985); Coughlin, S.R. et al., Science 243:1191-1194 (1989)). The absence of an ATP binding site argues that GRB-1 is most likely not a phospholipid kinase. GRB-1 exhibits 97% sequence identity with murine and bovine p85. Hence, GRB-1 is the human counterpart of p85. Recombinant p85 is able to bind to the activated PDGFR or EGFR, but does not itself contain intrinsic PI3 kinase activity. p85, however, is found associated with a 110 kDa tyrosine phosphorylated protein which may be the catalytic subunit of the PI3 Kinase. While the exact relationship between PI3 kinase and p85 is not known, overexpression of p85 modulates the interaction between PI3 kinase and the PDGFR. p85 could function as a regulatory subunit or as a bridge between activated receptors and the PI3 kinase.

EXAMPLE VI

The Tyrosine Phosphorylated Carboxy-terminus of the EGF Receptor is a Binding Site for GAP and PLC-gamma

The studies described below confirm that binding of PLC-gamma and a fusion protein containing the SH2 and SH3 domains of GAP (trpE/GAP SH2) are specifically controlled by autophosphorylation of the EGFR. The results show that phosphorylation of PLC-gamma actually reduces its association with the EGFR. Evidence is presented demonstrating that both PLC-gamma and the trpE/GAP SH2 fusion protein bind

specifically to the tyrosine phosphorylated C-terminus of the EGFR. In sum, these results indicate that the SH2/SH3 domains interact directly with phosphotyrosine containing regions of the EGF receptor.

A. Materials and Methods

1. Cell lines, mutant receptors and fusion proteins

The cell lines CD126 (Margolis, B.L. et al., J. Biol. Chem. 264: 10667-10671 (1989a), HER14, K721 (Honegger, A.M. et al., Cell 51: 199-209 (1987); Honegger, A.M. et al., Mol. Cell. Biol. 7:4567-4571 (1987)) were used as sources for wild-type EGF receptor, kinase-negative (kin^-) EGF receptor and C-terminal (C-terminal) truncated EGF receptor, respectively. The intracellular domain of the EGF receptor (EGFR-C) was purified from a baculovirus expression system (Hsu, C-.J. et al., Cell Growth Differ 1: 191-200 (1990)) (Figure 9A). 3TP1, a cell line which overexpresses transfected PLC-gamma cDNA but has no EGF receptor was used as a source of PLC-gamma (Margolis, B. et al., Science 248: 607-610 (1990b)).

The preparation of trpE fusion proteins containing the GAP SH2 domain (GAP residues 171-448, Figure 9B) has been described by Moran, M.F. et al., Proc. Natl. Acad. Sci. USA 87: 8622-8626 (1990). Bacterial lysates containing trpE/GAP SH2 fusion proteins were prepared by resuspending 1 g of bacteria in 3 ml of 50 mM Tris pH 7.5, 0.5 mM EDTA, 0.1 mM PMSF. After incubation at 4°C in 1 mg/ml lysozyme and 0.2% NP-40, cells were sonicated 5 times for 5 seconds, and the lysate was clarified by centrifugation for 30 min at 10,000 g. Bacterial lysates were diluted 1:100 in the 1% Triton lysis buffer with proteinase and phosphatase inhibitors as described above and were precleared with protein A-Sepharose.

2. Antibodies, immunoprecipitation and immunoblotting

The following anti-EGFR antibodies (Figure 9A) were used: (a) mAb108, a monoclonal antibody directed against domain III of the extracellular domain (Lax, I. et al., EMBO J. 8: 421-427 (1989)); (b) antipeptide antibody RK2 specific for residues 984-996; (c) antipeptide antibody C specific for residues 1176-1186; and (d) antipeptide antibody F, specific

for residues 656-676. For immunoprecipitating the trpE fusion proteins, a mouse monoclonal antibody against trpE (Oncogene Science) bound to agarose linked anti-mouse IgG (Sigma) was utilized. For immunoblotting, a polyclonal rabbit antibody against trpE was used (Moran, M.F. et al., Proc. Natl. Acad. Sci. USA 87: 8622-8626 (1990)). PLC-gamma was immunoblotted and immunoprecipitated with a polyclonal rabbit anti-peptide antibody described previously (Margolis, B. et al., Cell 57: 1101-1107 (1989b)).

The techniques used are described in several references from the present inventors' laboratory (Margolis, B.L. et al., J. Biol. Chem. 264: 10667-10671 (1989); Cell 57:1101-1107 (1989)). Unstimulated cells were grown to confluence in Dulbecco's Modified Eagle Medium with 10% calf serum and starved overnight in 1% fetal calf serum prior to lysis in a 1% Triton X-100 lysis buffer containing proteinase and phosphatase inhibitors. EGF receptors were immunoprecipitated utilizing antibodies bound to protein A-Sepharose. After washing the receptor material with HNTG (20 mM Hepes, pH 7.5, 150 mM NaCl, 0.1% Triton X-100 and 10% glycerol), autophosphorylation was induced by the addition of 5mM $MnCl_2$ and 30 μM ATP. Controls were incubated with Mn^{2+} only. After further washes with HNTG, lysate containing either PLC-gamma (from 3TP1 cells) or the bacterial fusion proteins was added. After allowing binding to proceed for 90 min, three further washes with HNTG were performed and samples were run on an SDS gel and immunoblotted.

3. Cyanogen bromide (CNBr) cleavage

EGFR-C was phosphorylated at 4°C with $MnCl_2$ and ATP sometimes in the presence of [gamma-^{32}P]ATP (NEN/Dupont, 6000 Ci/mmol). The receptor preparation was then resuspended in 20 mM HEPES, pH 7.5, with 100 μg BSA and concentrated in a Centricon 10 (Amicon) to 50 μl. Then 240 μl 88% formic acid was added with two grains of CNBr and the samples were stored under nitrogen in the dark for 14 h at room temperature. Samples were dried and washed three times with water in a Speed-Vac (Savant) and then resuspended in 1% Triton lysis buffer.

B. RESULTS

A comparison was performed of the binding of PLC-gamma to wild-type and mutant EGFRs (Figure 9A). First, wild-type and mutant receptors from transfected NIH-3T3 cells were immunoprecipitated and some of the receptor immunoprecipitates were allowed to undergo in vitro autophosphorylation with ATP and Mn^{2+} (Margolis, B. et al., Mol. Cell. Biol. 10: 435-441 (1990a)). Then, lysates from NIH-3T3 cells which overexpress PLC-gamma (Margolis, B. et al., Science 248: 607-610 (1990b)) were added and binding allowed to proceed for 90 min. at 4°C. After washing the immunoprecipitates with HNTG, the amount of PLC-gamma bound was assessed by immunoblotting.

As illustrated in Figure 10, PLC-gamma bound only to the tyrosine phosphorylated wild-type receptor but not to the non-phosphorylated receptor.

To assess the importance of autophosphorylation, two studies with mutant receptors were then undertaken. First to be examined was the binding of PLC-gamma to a truncated EGF receptor missing 126 amino acids from the C-terminus (CD126, Figure 9A) and devoid of four major autophosphorylation sites (Downward, J. et al., Nature 311: 483-485 (1984)). This truncated receptor was autophosphorylated, probably at tyrosine 992 (Walton, G.M. et al., J. Biol. Chem. 265: 1750-1754 (1990)). However, despite this level of tyrosine autophosphorylation, the binding of PLC-gamma was markedly reduced compared to the full length receptor. Reduced association was also observed with CD63, a deletion mutant EGF receptor lacking 63 C-terminal residues containing two autophosphorylation sites. These results suggested a role for the receptor C-terminus in either binding or modulating the binding of PLC-gamma to the EGF receptor.

Figure 10 also demonstrates that PLC-gamma cannot bind to the kin^- mutant receptor. To explore the importance of autophosphorylation in this effect, the kin^- receptor was cross-phosphorylated with the CD126 receptor (Honegger, A.M. et al., Proc. Natl. Acad. Sci. USA 86:925-929 (1989)). This resulted in normalization of PLC-gamma binding to wild-type

levels. This suggested that phosphorylation of the kin⁻ receptor was sufficient to normalize binding to PLC-gamma.

To confirm that the kin⁻ receptor alone could bind PLC-gamma after phosphorylation, this receptor was cross-phosphorylated with a soluble, baculovirus-expressed EGFR cytoplasmic domain (EGFR-C) that does not bind to the mAb 108 (Figure 9A).

Although cross-phosphorylation was not as strong as with the CD126 mutant, tyrosine phosphorylation of the K721A mutant and binding of PLC-gamma were clearly detected. This finding confirms that tyrosine phosphorylation of the EGFR promotes binding of PLC-gamma.

The role of PLC-gamma tyrosine phosphorylation in the interaction between wild-type EGFR and PLC-gamma was examined. Tyrosine phosphorylated PLC-gamma could be dissociated from the EGFR more readily than non-phosphorylated PLC-gamma (Figure 11), suggesting a lower affinity of tyrosine phosphorylated PLC-gamma for the EGFR.

These findings were extended to examination of the binding of a fusion protein containing trpE/GAP SH2 domain (Figure 9B) to the baculovirus expressed EGFR-C. As with the full length EGFR and PLC-gamma, the trpE/GAP SH2 fusion protein domain bound only to the tyrosine phosphorylated EGFR-C (Figure 12A). The trpE protein alone did not bind to EGFR-C. Similarly, phosphorylated EGFR-C bound only to trpE/GAP SH2; however, non-specific binding of non-phosphorylated EGFR-C was high (Figure 12B). These results demonstrated that the binding site of the EGFR is situated in its intracellular domain.

In general, the trpE/GAP SH2 fusion protein bound with a higher stoichiometry to full length EGFR than did PLC-gamma. However, the fusion protein was not tyrosine phosphorylated by the EGFR. The trpE/GAP SH2 protein much better to the phosphorylated full length receptor compared to the CD126 deletion mutant (Figure 13A). As shown in Figure 13B, cross-phosphorylation of the kin⁻ full length EGF receptor by the EGFR-C allowed it to bind the trpE/GAP SH2 protein.

In control groups, the EGFR-C was shown not to enhance the binding to the CD126 receptor probably because this receptor was already maximally tyrosine phosphorylated (Figure 13A). Also, no binding was observed when EGFR-C was tested in the presence of mAb 108 immunoprecipitate from cells containing no EGF receptor (Figure 13B). This indicates that the effects of EGFR-C could not be attributed to non-specific binding of tyrosine phosphorylated EGFR-C to sepharose. These studies confirm the importance of autophosphorylation in mediating binding and show that for EGF receptor binding, the GAP SH2 domain behaves similarly to intact PLC-gamma.

The poor binding to the CD126 deletion mutant suggested that at least part of the binding site for the molecule was in the C-terminus. Yet an effect, possibly allosteric, of this deletion on the overall conformation of the receptor could not be excluded. Therefore, the binding of PLC-gamma and trpE/GAP SH2 to a C-terminal fragment of the EGFR was examined. In the EGFR, the most C-terminal methionine residue is found at position 983; CNBr cleavage therefore generates a 203 amino acid fragment which contains all the known autophosphorylation sites. This protein fragment is recognized by an antibody specific for the EGFR C-terminus, anti-C (Figure 9A).

When this C-terminal fragment was specifically immunoprecipitated and tyrosine phosphorylated, it bound PLC-gamma and the trpE/GAP SH2 fusion protein (Figure 14). CNBr cleavage was complete; no full-length EGFR-C could be detected after proteolysis that could account for the binding. Again, no binding was seen to the non-phosphorylated C-terminal CNBr fragment. CNBr cleavage of EGFR-C also generated a 97 amino acid N-terminal peptide identified by antibody F (Figure 9A, EGFR residues 645-742). This fragment, immunoprecipitated by antibody F, did not bind trpE/GAP SH2. Additionally, EGFR-C was autophosphorylated with [gamma-^{32}P]ATP and a ^{32}P-labeled CNBr C-terminal fragment was generated. As shown in Figure 15, this fragment bound to the trpE/GAP SH2 fusion protein but not to trpE. In total, these findings demonstrate that direct binding to the tyrosine phosphorylated C-terminus contributes

at least in part to the specific binding of SH2 and SH3 domain proteins to the EGFR.

C. Discussion

When taken together, the above findings and several additional lines of evidence argue strongly that the phosphotyrosine residues are part of the actual binding site of the EGFR for SH2 domains. First, $P47^{gag-crk}$ was found to bind to nearly all phosphotyrosine-containing proteins in v-crk transformed cells (Matsuda, M. et al., Science 248: 1537-1539 (1990)). Second, mutations of two autophosphorylation sites on the PDGF receptor greatly decreased the binding of GAP (Kazlauskas, A. et al., Science 247: 1578-1581 (1990)). Finally, the results presented above demonstrate specific binding to the C-terminus of the EGFR only when phosphotyrosine is present. Thus, it is concluded that the phosphotyrosine residues either comprise a part of the binding site or locally alter the conformation of this region, allowing binding. It is unlikely that phosphotyrosine alone constitutes the binding site. For example, phosphotyrosine alone cannot interfere with the binding of $P47^{gag-crk}$ to phosphotyrosine-containing proteins (Matsuda et al., supra). Additionally, PLC-gamma does not bind to activated all molecules that contain phosphotyrosine residues, such as the CSF-1 receptor (Downing, J.R. et al., EMBO J. 8:3345-3350 (1989)). Similarly, the binding of PLC-gamma to PDGFR does not appear to be identical to GAP binding; different SH2 and SH3 domain-containing proteins may have different binding specificities (Kazlauskas et al., supra).

The references cited above are all incorporated by reference herein, whether specifically incorporated or not.

Having now fully described this invention, it will be appreciated by those skilled in the art that the same can be performed within a wide range of equivalent parameters, concentrations, and conditions without departing from the spirit and scope of the invention and without undue experimentation.

While this invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications. This application is intended to cover any variations, uses, or adaptations of the inventions following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth as follows in the scope of the appended claims. It is to be understood that the phraseology or terminology employed herein is for the purpose of description and not of limitation.

WHAT IS CLAIMED IS:

1. A method for detecting expression in a cell of a protein which is capable of binding to a tyrosine-phosphorylated polypeptide portion of a receptor tyrosine kinase molecule, said method comprising:

(a) contacting said cell, an extract thereof, a lysate thereof, or a supernatant thereof with a solid phase carrier, causing the binding of said protein to said carrier;

(b) incubating said carrier-bound protein with said tyrosine-phosphorylated polypeptide, allowing said polypeptide to bind to said carrier-bound protein;

(c) removing materials not bound to said carrier; and

(d) detecting the presence or measuring the amount of said tyrosine-phosphorylated polypeptide bound to said carrier,

thereby detecting the expression of said protein.

2. A method according to claim 1 wherein said receptor tyrosine kinase is selected from the group consisting of epidermal growth factor receptor, platelet-derived growth factor receptor, fibroblast growth factor receptor, colony stimulating factor-1 receptor and insulin receptor.

3. A method according to claim 1 wherein said protein is of human origin.

4. A method according to claim 1 wherein said phosphorylated polypeptide is detectably labeled.

5. A method according to claim 4 wherein said detectable label is a radiolabel.

6. A method according to claim 5 wherein said radio label is ^{32}P.

7. A method for mapping to a human chromosome a gene encoding a protein which is capable of binding to a tyrosine-phosphorylated polypeptide portion of a receptor tyrosine kinase molecule, said method comprising:

(a) infecting bacterial cells with a human gene expression library;

(b) detecting a clone expressing said protein using a method according to claim 1;

(c) sequencing the DNA of said clone; and

(d) mapping said sequence to a human chromosome.

8. A polypeptide probe useful in the detection of the expression of a protein capable of binding to a tyrosine-phosphorylated polypeptide portion of a receptor tyrosine kinase, said probe comprising an amino acid sequence derived from the tyrosine-phosphorylated portion of said receptor molecule, or a functional derivative thereof, wherein said sequence has at least one phosphorylated tyrosine residue and lacks the tyrosine kinase portion of said receptor, and wherein said probe is detectably labeled.

9. A probe according to claim 8 having between one and five phosphorylated tyrosine residues.

10. A probe according to claim 8 between 25 and 250 amino acid residues.

11. A probe according to claim 8 and wherein the detectable label is a radioactive phosphorous atom of said phosphorylated tyrosine residue.

12. A method for preparing a probe according to claim 8, comprising:

(a) providing said receptor, or a genetically engineered receptor-like derivative in substantially pure form, wherein said receptor or said receptor-like derivative has both a tyrosine kinase domain and a tyrosine-phosphorylated domain, said tyrosine-phosphorylated domain including at least one tyrosine residue capable of being phosphorylated by said tyrosine kinase;

(b) incubating said receptor or receptor-like derivative with [gamma-^{32}P]adenosine triphosphate under conditions permitting phosphorylation of said tyrosine residue, causing phosphorylation of said tyrosine residue;

thereby producing said probe.

13. A method according to claim 12, comprising, after step (b),

(c) additionally treating said phosphorylated receptor molecule with an agent capable of cleaving said molecule between the tyrosine kinase domain and the tyrosine-phosphorylated domain.

14. A method according to claim 13 wherein said agent is cyanogen bromide.

15. A method according to claim 12, wherein said receptor-like derivative is a polypeptide encoded by a DNA molecule comprising a DNA sequence encoding tyrosine kinase, linked to a DNA sequence encoding a selective enzymatic cleavage site, linked to a DNA sequence encoding said carboxyterminal domain, and wherein said agent is an enzyme capable of cleaving at said cleavage site.

16. A method for purifying from a complex mixture a protein which is capable of binding to a tyrosine-phosphorylated polypeptide portion of a receptor tyrosine kinase molecule, said method comprising:

(a) contacting said complex mixture with a solid phase carrier to which a probe according to claim 8 is bound, allowing said protein to bind to said probe;

(b) removing materials not bound to said carrier; and

(c) eluting said bound protein from said carrier, thereby purifying said protein.

17. A protein, GRB-1, having the amino acid sequence shown in Figure 4.

18. A protein, GRB-2, having the amino acid sequence shown in Figure 16.

19. A DNA molecule encoding a protein according to claim 17, or a functional derivative thereof, wherein, when said DNA molecule naturally occurs, it is substantially free of the nucleotide sequences with which it is natively associated.

20. A DNA molecule encoding a protein according to claim 18, or a functional derivative thereof, wherein, when said DNA molecule naturally occurs, it is substantially free of the nucleotide sequences with which it is natively associated.

21. A DNA molecule consisting essentially of the nucleotide sequence shown in Figure 4.

22. A DNA molecule encoding a protein according to claim 18, or encoding a functional derivative thereof.

FIG. 1

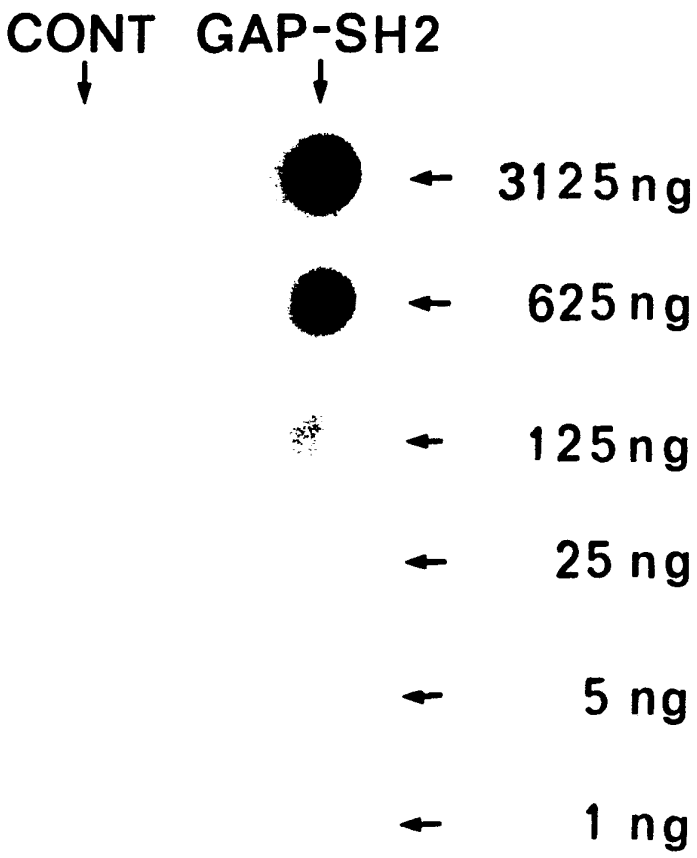

FIG. 2

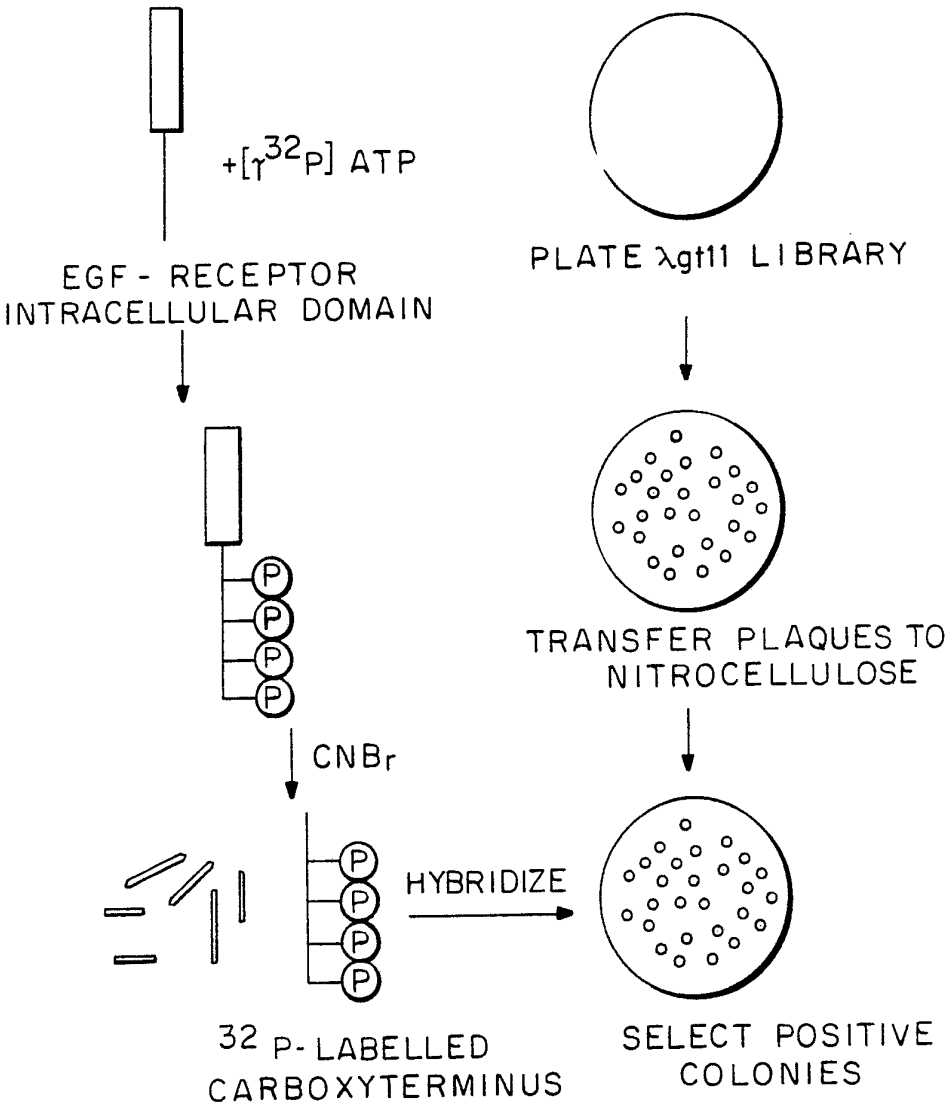
+[γ32P] ATP
EGF - RECEPTOR INTRACELLULAR DOMAIN
PLATE λgt11 LIBRARY
TRANSFER PLAQUES TO NITROCELLULOSE
CNBr
HYBRIDIZE
32 P-LABELLED CARBOXYTERMINUS
SELECT POSITIVE COLONIES

FIG. 3A

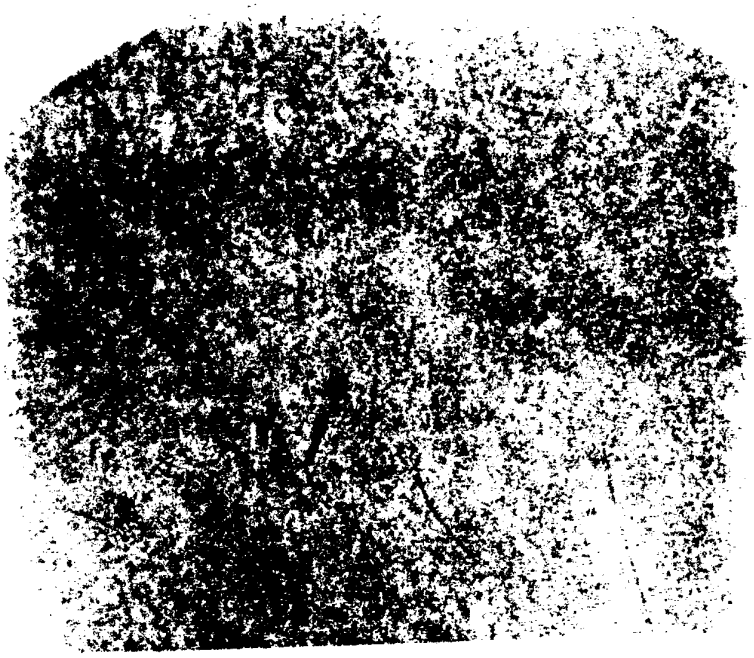

FIG. 3B



FIG. 4A

```
                                                                    TACAACCAGGCTCAACTGTTGCATGGTAGCAGATTTGCAAACATGAGTGCTGAGGGGTAC
    1 ----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+ 60
                                                                    ATGTTGGTCCGAGTTGACAACGTACCATCGTCTAAACGTTTGTACTCACGACTCCCCATG

                                                                                                                  M  S  A  E  G  Y  -

                                                                    CAGTACAGAGCGCCTGTATGATTATAAAAGGAAAGAGAAGAAGATATTGACTTGCACTTG
   61 ----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+ 120
                                                                    GTCATGTCTCGCGACATACTAATATTTTTCCTTTCTCTTCTTCTATAACTGAACGTGAAC

                                                                    Q  Y  R  A  L  Y  D  Y  K  K  K  E  E  R  E  E  D  I  D  L  H  L

                                                                    GGTGACATATTGACTGTGAATAAAGGGTCCTTAGTAGCTCTTGGATTCAGTGATGGACAG
  121 ----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+ 180
                                                                    CCACTGTATAACTGACACTTATTTCCCAGGAATCATCGAGAACCTAAGTCACTACCTGTC

                                                                    G  D  I  L  T  V  N  K  G  S  L  V  A  L  G  F  S  D  G  Q  -

                                                                    GAAGCCAGGCCTGAAGAAATTGGCTGGTTAAATGGCTATAATGAAACCACAGGGGAAAGG
  181 ----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+ 240
                                                                    CTTCGGTCCGGACTTCTTTAACCGACCAATTTACCGATATTACTTTGGTGTCCCCTTTCC

                                                                    E  A  R  P  E  E  I  G  W  L  N  G  Y  N  E  T  T  G  E  R

                                                                    GGGGACTTTCCGGGAACTTACGTAGAATATATTGGAAGGAAAAAAATCTCGCCTCCCACA
  241 ----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+ 300
                                                                    CCCCTGAAAGGCCCTTGAATGCATCTTATATAACCTTCCTTTTTTTTAGAGCGGAGGGTGT

                                                                    G  D  F  P  G  T  Y  V  E  Y  I  G  R  K  K  I  S  P  P  T

                                                                    CCAAAGCCCCGGCCACCTCGGCCTCTTCCTGTTGCACCAGGTTCTTCGAAAACTGAAGCA
  301 ----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+ 360
                                                                    GGTTTCGGGGCCGGTGGAGCCGGAGAAGGACAACGTGGTCCAAGAAGCTTTTGACTTCGT

                                                                    P  K  P  R  P  P  R  P  L  P  V  A  P  G  S  S  K  T  E  A  -
```

SUBSTITUTE SHEET

FIG. 4B

```
          GATGTTGAACAACAAGCTTTGACTCTCCCGGATCTTGCAGAGCAGTTTGCCCCTCCTGAC
      361 ---------+---------+---------+---------+---------+---------+ 420
          CTACAACTTGTTGTTCGAAACTGAGAGGGCCTAGAACGTCTCGTCAAACGGGGAGGACTG

  a       D  V  E  Q  Q  A  L  T  L  P  D  L  A  E  Q  F  A  P  P  D  -

          ATTGCCCCGCCTCTTCTTATCAAGCTCGTGGAAGCCATTGAAAAGAAAGGTCTGGAATGT
      421 ---------+---------+---------+---------+---------+---------+ 480
          TAACGGGGCGGAGAAGAATAGTTCGAGCACCTTCGGTAACTTTTCTTTCCAGACCTTACA

  a       I  A  P  P  L  L  I  K  L  V  E  A  I  E  K  K  G  L  E  C  -

      481 TCAACTCTATACAGAACACAGAGCTCCAGCAACCTGGCAGAATTACGACAGCTTCTTGAT
          ---------+---------+---------+---------+---------+---------+ 540
          AGTTGAGATATGTCTTGTGTCTCGAGGTCGTTGGACCGTCTTAATGCTGTCGAAGAACTA

  a       S  T  L  Y  R  T  Q  S  S  S  N  L  A  E  L  R  Q  L  L  D  -

      541 TGTGATACACCCTCCGTGGACTTGGAAATGATCGATGTGCACGTTTTGGCTGACGCTTTC
          ---------+---------+---------+---------+---------+---------+ 600
          ACACTATGTGGGAGGCACCTGAACCTTTACTAGCTACACGTGCAAAACCGACTGCGAAAG

  a       C  D  T  P  S  V  D  L  E  M  I  D  V  H  V  L  A  D  A  F  -

      601 AAACGCTATCTCCTGGACTTACCAAATCCTGTCATTCCAGCAGCCGTTTACAGTGAAATG
          ---------+---------+---------+---------+---------+---------+ 660
          TTTGCGATAGAGGACCTGAATGGTTTAGGACAGTAAGGTCGTCGGCAAATGTCACTTTAC

  a       K  R  Y  L  L  D  L  P  N  P  V  I  P  A  A  V  Y  S  E  M  -

      661 ATTTCTTTAGCTCCAGAAGTACAAAGCTCCGAAGAATATATTCAGCTATTGAAGAAGCTT
          ---------+---------+---------+---------+---------+---------+ 720
          TAAAGAAATCGAGGTCTTCATGTTTCGAGGCTTCTTATATAAGTCGATAACTTCTTCGAA

  a       I  S  L  A  P  E  V  Q  S  S  E  E  Y  I  Q  L  L  K  K  L  -
```

FIG. 4C

```
                ATTAGGTCGCCTAGCATACCTCATCAGTATTGGCTTACGCTTCAGTATTTGTTAAAACAT
          721   ---------+---------+---------+---------+---------+---------+  780
                TAATCCAGCGGATCGTATGGAGTAGTCATAACCGAATGCGAAGTCATAAACAATTTTGTA

a                I  R  S  P  S  I  P  H  Q  Y  W  L  T  L  Q  Y  L  L  K  H

                TTCTTCAAGCTCTCTCAAACCTCCAGCAAAAATCTGTTGAATGCAAGAGTACTCTCTGAA
          781   ---------+---------+---------+---------+---------+---------+  840
                AAGAAGTTCGAGAGAGTTTGGAGGTCGTTTTTAGACAACTTACGTTCTCATGAGAGACTT

a                F  F  K  L  S  Q  T  S  S  K  N  L  L  N  A  R  V  L  S  E  -

                ATTTTCAGCCCTATGCTTTTCAGATTCTCAGCAGCCAGCTCTGATAATACTGAAAACCTC
          841   ---------+---------+---------+---------+---------+---------+  900
                TAAAAGTCGGGATACGAAAAGTCTAAGAGTCGTCGGTCGAGACTATTATGACTTTTGGAG

a                I  F  S  P  M  L  F  R  F  S  A  A  S  S  D  N  T  E  N  L  -

                ATAAAAGTTATAGAAATTTTAATCTCAACTGAATGGAATGAACGACAGCCTGCACCAGCA
          901   ---------+---------+---------+---------+---------+---------+  960
                TATTTTGAATATGTTTAAAATTAGAGTTGACTTACCTTACTTGCTGTCGGACGTGGTCGT

a                I  K  V  I  E  I  L  I  S  T  E  W  N  E  R  Q  P  A  P  A

                CTGCCTCCTAAACCACCAAAACCTACTACTGTAGCCAACAACGGTATGAATAACAATATG
          961   ---------+---------+---------+---------+---------+---------+ 1020
                GACGGAGGATTTGGTGGTTTTGGATGATGACATCGGTTGTTGCCATACTTATTGTTATAC

a                L  P  P  K  P  P  K  P  T  T  V  A  N  N  G  M  N  N  N  M  -

                TCCTTACAAAATGCTGAATGGTACTGGGGAGATATCTCGAGGGAAGAAGTGAATGAAAAA
         1021   ---------+---------+---------+---------+---------+---------+ 1080
                AGGAATGTTTTACGACTTACCATGACCCCTCTATAGAGCTCCCTTCTTCACTTACTTTTT

a                S  L  Q  N  A  E  W  Y  W  G  D  I  S  R  E  E  V  N  E  K  -
```

SUBSTITUTE SHEET

FIG. 4D

```
          CTTCGAGATACAGCAGACGGGACCTTTTTGGTACGAGATGCGTCTACTAAAATGCATGGT
     1081 ---------+---------+---------+---------+---------+---------+
          GAAGCTCTATGTCGTCTGCCCTGGAAAAACCATGCTCTACGCAGATGATTTTACGTACCA 1140

a              L  R  D  T  A  D  G  T  F  L  V  R  D  A  S  T  K  M  H  G  -

          GATTATACTCTTACACTAAGGAAAGGGGGAAATAACAAATTAATCAAAATATTTCATCGA
     1141 ---------+---------+---------+---------+---------+---------+---------+ 1200
          CTAATATGAGAATGTGATTCCTTTCCCCCTTTATTGTTTAATTAGTTTTATAAAGTAGCT

a              D  Y  T  L  T  L  R  K  G  G  N  N  K  L  I  K  I  F  H  R  -

          GATGGGAAATATGGCTTCTCTGACCCATTAACCTTCAGTTCTGTGGTTGAATTAATAAAC
     1201 ---------+---------+---------+---------+---------+---------+---------+ 1260
          CTACCCTTTATACCGAAGAGACTGGGTAATTGGAAGTCAAGACACCAACTTAATTATTTG

a              D  G  K  Y  G  F  S  D  P  L  T  F  S  S  V  V  E  L  I  N  -

          CACTACCGGAATGAATCTCTAGCTCAGTATAATCCCAAATTGGATGTGAAATTACTTTAT
     1261 ---------+---------+---------+---------+---------+---------+---------+ 1320
          GTGATGGCCTTACTTAGAGATCGAGTCATATTAGGGTTTAACCTACACTTTAATGAAATA

a              H  Y  R  N  E  S  L  A  Q  Y  N  P  K  L  D  V  K  L  L  Y  -

     1321 CCAGTATCCAAATACCAACAGGATCAAGTTGTCAAAGAAGATAATATTGAAGCTGTAGGG
          ---------+---------+---------+---------+---------+---------+---------+ 1380
          GGTCATAGGTTTATGGTTGTCCTAGTTCAACAGTTTCTTCTATTATAACTTCGACATCCC

a              P  V  S  K  Y  Q  Q  D  Q  V  V  K  E  D  N  I  E  A  V  G  -

          AAAAAATTACATGAATATAACACTCAGTTTCAAGAAAAAAGTCGAGAATATGATAGATTA
     1381 ---------+---------+---------+---------+---------+---------+---------+ 1440
          TTTTTTAATGTACTTATATTGTGAGTCAAAGTTCTTTTTTCAGCTCTTATACTATCTAAT

a              K  K  L  H  E  Y  N  T  Q  F  Q  E  K  S  R  E  Y  D  R  I  -
```

FIG. 4E

```
                                TATGAAGAATATACCCGCACATCCCAGGAAATCCAAATGAAAAGGACAGCTATTGAAGCA
1441 ----------+----------+----------+----------+----------+----------+----------+ 1500
                ATACTTCTTATATGGGCGTGTAGGGTCCTTTAGGTTTACTTTTCCTGTCGATAACTTCGT

a                Y  E  E  Y  T  R  T  S  Q  E  I  Q  M  K  K  R  R  T  A  I  E  E  A  -

                TTTAATGAAACCATAAAAATATTTGAAGAACAGTGCCAGACCCAAGAGCGGTACAGCAAA
1501 ----------+----------+----------+----------+----------+----------+----------+ 1560
                AAATTACTTTGGTATTTTTATAAACTTCTTGTCACGGTCTGGGTTCTCGCCATGTCGTTT

a                F  N  E  T  I  K  I  F  E  E  Q  C  Q  T  Q  E  R  Y  S  K

                GAATACATAGAAAAGTTTAAACGTGAAGGCAATGAGAAAGAAATACAAAGGATTATGCAT
1561 ----------+----------+----------+----------+----------+----------+----------+ 1620
                CTTATGTATCTTTTCAAATTTGCACTTCCGTTACTCTTTCTTTATGTTTCCTAATACGTA

a                E  Y  I  E  K  F  K  R  E  G  N  E  K  E  I  Q  R  I  M  H  -

                AATTATGATAAGTTGAAGTCTCGAATCAGTGAAATTATTGACAGTAGAAGAAGATTGGAA
1621 ----------+----------+----------+----------+----------+----------+----------+ 1680
                TTAATACTATTCAACTTCAGAGCTTAGTCACTTTAATAACTGTCATCTTCTTCTAACCTT

a                N  Y  D  K  L  K  S  R  I  S  E  I  I  D  S  R  R  R  L  E  -

                GAAGACTTGAAGAAGCAGGCAGCTGAGTATCGAGAAATTGACAAACGTATGAACAGCATT
1681 ----------+----------+----------+----------+----------+----------+----------+ 1740
                CTTCTGAACTTCTTCGTCCGTCGACTCATAGCTCTTTAACTGTTTGCATACTTGTCGTAA

a                E  D  L  K  K  Q  A  A  E  Y  R  E  I  D  K  R  M  N  S  I  -

                AAACCAGACCTTATCCAGCTGAGAAAGACGAGAGACCAATACTTGATGTGGTTGACTCAA
1741 ----------+----------+----------+----------+----------+----------+----------+ 1800
                ATTGGTCTGGAATAGGTCGACTCTTTCTGCTCTCTGGTTATGAACTACACCAACTGAGTT

                K  P  D  L  I  Q  L  R  K  K  T  R  D  Q  Y  L  M  W  L  T  Q  -
```

SUBSTITUTE SHEET

FIG. 4F

```
                AAAGGTGTTCGGCAAAAGAAGTTGAACGAGTGGTTGGGCAATGAAAACACTGAAGACCAA
      1801      ---------+---------+---------+---------+---------+---------+  1860
                TTTCCACAAGCCGTTTTCTTCAACTTGCTCACCAACCCGTTACTTTTGTGACTTCTGGTT

a                K  G  V  R  Q  K  K  L  N  E  W  L  G  N  E  N  T  E  D  Q

                TATTCACTGGTGGAAGATGATGAAGATTTGCCCCATCATGATGAGAAGACATGGAATGTT
      1861      ---------+---------+---------+---------+---------+---------+  1920
                ATAAGTGACCACCTTCTACTACTTCTAAACGGGGTAGTACTACTCTTCTGTACCTTACAA

a                Y  S  L  V  E  D  D  E  D  L  P  H  H  D  E  K  T  W  N  V  -

                GGAAGCAGCAACCGAAACAAAGCTGAAAACCTGTTGCGAGGGAAGCGAGATGGCACTTTT
      1921      ---------+---------+---------+---------+---------+---------+  1980
                CCTTCGTCGTTGGCTTTGTTTCGACTTTTGGACAACGCTCCCTTCGCTCTACCGTGAAAA

a                G  S  S  N  R  N  K  A  E  N  L  L  R  G  K  R  D  G  T  F  -

                CTTGTCCGGGAGAGCAGTAAACAGGGCTGCTATGCCTGCTCTGTAGTGGTGGACGGCGAA
      1981      ---------+---------+---------+---------+---------+---------+  2040
                GAACAGGCCCTCTCGTCATTTGTCCCGACGATACGGACGAGACATCACCACCTGCCGCTT

a                L  V  R  E  S  S  K  Q  G  C  Y  A  C  S  V  V  V  D  G  E  -

                GTAAAGCATTGTGTCATAAACAAAACAGCAACTGGCTATGGCTTTGCCGAGCCCTATAAC
      2041      ---------+---------+---------+---------+---------+---------+  2100
                CATTTCGTAACACAGTATTTGTTTTGTCGTTGACCGATACCGAAACGGCTCGGGATATTG

a                V  K  H  C  V  I  N  K  T  A  T  G  Y  G  F  A  E  P  Y  N  -

                TTGTACAGCTCTCTGAAAGAACTGGTGCTACATTACCAACACACCTCCCTTGTGCAGCAC
      2101      ---------+---------+---------+---------+---------+---------+  2160
                AACATGTCGAGAGACTTTCTTGACCACGATGTAATGGTTGTGTGGAGGGAACACGTCGTG

a                L  Y  S  S  L  K  E  L  V  L  H  Y  Q  H  T  S  L  V  Q  H  -
```

FIG. 4G

```
      AACGACTCCCTCAATGTCACACTAGCCTACCCAGTATATGCACAGCAGAGGCGATGAAGC
2161  ---------+---------+---------+---------+---------+---------+ 2220
      TTGCTGAGGGAGTTACAGTGTGATCGGATGGGTCATATACGTGTCGTCTCCGCTACTTCG

a         N  D  S  L  N  V  T  L  A  Y  P  V  Y  A  Q  Q  R  R     -

      GCTTACTCTTTGATCCTTCTCCTGAAGTTCAGCCACCCTGAGGCCTCTGGAAAGCAAAGG
2221  ---------+---------+---------+---------+---------+---------+ 2280
      CGAATGAGAAACTAGGAAGAGGACTTCAAGTCGGTGGGACTCCGGAGACCTTTCGTTTCC
a

      GCTCCTCTCCAGTCTGATCTGTGAATTGAGCTGCAGAAACGAAGCCATCTTTCTTTGGAT
2281  ---------+---------+---------+---------+---------+---------+ 2340
      CGAGGAGAGGTCAGACTAGACACTTAACTCGACGTCTTTGCTTCGGTAGAAAGAAACCTA
a

      GGGACTAGAGCTTTCTTTCACAAAAAAGAAGTAGGGGAAGACATGCAGCCTAAGGCTGTA
2341  ---------+---------+---------+---------+---------+---------+ 2400
      CCCTGATCTCGAAAGAAAGTGTTTTTTCTTCATCCCCTTCTGTACGTCGGATTCCGACAT
a

      TGATGACCACACGTTCCTAAGCTGGAGTGCTTATCCCTTCTTTTTCTTTTTTTCTTTGGT
2401  ---------+---------+---------+---------+---------+---------+ 2460
      ACTACTGGTGTGCAAGGATTCGACCTCACGAATAGGGAAGAAAAAGAAAAAAAGAAACCA
a

      TTAATTTAAAGCCACAACCACATACAACACAAAGAGAAAAAGAAATGCAAAAATCTCTGC
2461  ---------+---------+---------+---------+---------+---------+ 2520
      AATTAAATTTCGGTGTTGGTGTATGTTGTGTTTCTCTTTTTCTTTACGTTTTTAGAGACG
a

      GTGCAGGGACAAAGAGGCCTTTAACCATGGTGCTTGTTAATGCTTTCTGAAGCTTTACCA
2521  ---------+---------+---------+---------+---------+---------+ 2580
      CACGTCCCTGTTTCTCCGGAAATTGGTACCACGAACAATTACGAAAGACTTCGAAATGGT
```

FIG. 4H

```
                GCTGAAAGTTGGGACTCTGGAGAGCGGAGGAGAGAGAGGCAGAAGAACCCTGGCCTGAGA
          2581  ---------+---------+---------+---------+---------+---------+  2640
                CGACTTTCAACCCTGAGACCTCTCGCCTCCTCTCTCTCCGTCTTCTTGGGACCGGACTCT
a

                AGGTTTGGTCCAGCCTGGTTTAGCCTGGATGTTGCTGTGCACGGTGGACCCAGACACATC
          2641  ---------+---------+---------+---------+---------+---------+  2700
                TCCAAACCAGGTCGGACCAAATCGGACCTACAACGACACGTGCCACCTGGGTCTGTGTAG
a

                GCACTGTGGATTATTTCATTTTGTAACAAATGAACGATATGTAGCAGAAAGGCACGTCCA
          2701  ---------+---------+---------+---------+---------+---------+  2760
                CGTGACACCTAATAAAGTAAAACATTGTTTACTTGCTATACATCGTCTTTCCGTGCAGGT
a

                CTCACAAGGGACGCTTTGGGAGAATGTCAGTTCATGTATGTTCAGAAGAAATTCTGTCAT
          2761   ---------+---------+---------+---------+---------+---------+  2820
                GAGTGTTCCCTGCGAAACCCTCTTACAGTCAAGTACATACAAGTCTTCTTTAAGACAGTA
a

                AGAAAGTGCCAGAAAGTGTTTAACTTGTCAAAAAACAAAAACCCAGCAACAGAAAAATGG
          2821  ---------+---------+---------+---------+---------+---------+  2880
                TCTTTCACGGTCTTTCACAAATTGAACAGTTTTTTGTTTTTGGGTCGTTGTCTTTTTACC
a

                AGTTTGGAAAACAGGACTTAAAATGACATTCAGTATATAAAATATGTACATAATATTGGA
          2881  ---------+---------+---------+---------+---------+---------+  2940
                TCAAACCTTTTGTCCTGAATTTTACTGTAAGTCATATATTTTATACATGTATTATAACCT
a

                TGACTAACTATCAAATAGATGGATTTGTATCAATACCAAATAGCTTCTGTTTTGTTTTGC
          2941  ---------+---------+---------+---------+---------+---------+  3000
                ACTGATTGATAGTTTATCTACCTAAACATAGTTATGGTTTATCGAAGACAAAACAAAACG
```

SUBSTITUTE SHEET

FIG. 4I

```
                TGAAGGCTAAATTCACAGCGCTATGCAATTCTTAATTTTCATTAAGTTGTTATTTCAGTT
3001 ---------+---------+---------+---------+---------+---------+---------+ 3060
                ACTTCCGATTTAAGTGTCGCGATACGTTAAGAATTAAAAGTAATTCAACAATAAAGTCAA
a

                TTAAATGTACCTTCAGAATAAGCTTCCCCACCCCAGTTTTTGTTGCTTGAAAATATTGTT
3061 ---------+---------+---------+---------+---------+---------+---------+ 3120
                AATTTACATGGAAGTCTTATTCGAAGGGGTGGGGTCAAAAACAACGAACTTTTATAACAA
a

                GTCCCGGATTTTTGTTAATATTCATTTTTGTTATCCTTTTTTAAAAATAAATGTACAGGA
3121 ---------+---------+---------+---------+---------+---------+---------+ 3180
                CAGGGCCTAAAAAGAATTATAAGTAAAAACAATAGGAAAAAATTTTTATTTACATGTCCT
a

                TGCCAGTAAAAAAAAAAATGGCTTCAGAATTAAAACTATGAAATATTTTACAGTTTTTCT
3181 ---------+---------+---------+---------+---------+---------+---------+ 3240
                ACGGTCATTTTTTTTTTTTACCGAAGTCTTAATTTTGATACTTTATAAAATGTCAAAAAGA
a

                TGTACAGAGTACTTGCTGTTAGCCCAAGGTTAAAAAGTTCATAACAGATTTTTTTTGGAC
3241 ---------+---------+---------+---------+---------+---------+---------+ 3300
                ACATGTCTCATGAACGACAATCGGGTTCCAATTTTTCAAGTATTGTCTAAAAAAAACCTG
a

                TGTTTTGTTGGGCAGTGCCTGATAAGCTTCAAAGCTGCTTTATTCAATAAAAAAAAAACC
3301 ---------+---------+---------+---------+---------+---------+---------+ 3360
                ACAAAACAACCCGTCACGGACTATTCGAAGTTTCGACGAAATAAGTTATTTTTTTTTTGG
a

                CGAATTCACTGG
3361 ----------+-- 3372
                GCTTAAGTGACC
```

FIG. 5

```
GRB-1 N 333  WYWGDIS -- R EE--- VN E- KL RDTAD------ GTFLVRDST KM HGDY T LT LRK-----GG--NN LIKI
GRB-1 C 624  WNVGSSN -- R NK--- AE N- LL RGKRD------ GTFLVRESS K- QGCY A CS VVV-----DG--EV KHCV

c-src   150  WYFGKIT -- R RE--- SE RL LL NPENPR----- GTFLVRESE TT KGAY C LS VSDF-DNAKGLNVK HYKI
v-abl   248  WYHGPVS -- R NA--- AE YK KS SGIN------- GSFLVRESE SS PG-Q R -S ISLRYE---G-RVY HYRI
PLC N   550  WFHGKLG AG R DGRHI AE R- LL TEYCIETGAPD GSFLVRESE TF VGDY T LS F--WR-N--G-KVQ HCRI
PLC C   668  WYHASLT -- R AQ--- AE H- ML MRVPRD----- GAFLVRKRN -E PNSY A IS F--RAE---G-KIK HCRV
GAP N   178  WYHGKLD -- R TI--- AE E- RL RQAGKS----- GSYLIRESD RR PGSF V LS FRSQM-N-V---VN HPRI
GAP C   348  WYHGKIS -- K QE--- AY N- LL MTVGQVC---- -SFLVRPSD NT PGDY S LY F-RTNENIQ--R-- -FKI
v-crk   248  WYWGRLS -- R GD--- AV S- LL QGQRH------ GTFLVRDSG SI PGDF V LS VSES---S---RVS HYIV

GRB-1 N 384   --FHRD G KYGFSDPLT---------------- F S SV V ELI N HY RNESLAQYNPKLDV- KL LY PVSK
GRB-1 C 672   INKTAT G -YGFAEPYNL--------------- Y S SL K ELV L HY QHTSLVQHNDSLNV- TL AY PVYA

c-src   207   RKLDSG G FYITSRTQ----------------- F S SL Q QLV A YY SKHADGLCH------ RL TN -VCP
v-abl   298   -NTASD G KLYVSSESR---------------- F N TL A ELV H HH STVADGLITT----- LH YP --AP
PLC N   611   HSRQDA G TPKFFLTDNLV-------------- F D SL Y DLI T HY QQVPLRCA-EFEM-- RL SL PV-P
PLC C   718   ---QQE G QTVMLGNSE----------------- F D SL V DLI S YY EKHPLYRK----M-- KL RY PI--
GAP N   230   --IAMC G DYYIGGRR----------------- F S SL S DLI G YY SHVSCLLKGE----- KL LY PVAP
GAP C   399   -CPTPN N QFMMGGRY----------------- Y N SI G DII D HY RKEQIVEG-YY---- -L KE PV-P
v-crk   298   NSLGPA G GRRAGGEGPFAPGLNPTRFLIGDNV F D SL P SLL E FY KIHYLDT--TT---- -L IE PV--

GRB-1    10  ALYDY KKEREE D IDLHLGDI LT VNK G SLVALGFSDPEARPEDIG WL NGYNETTGER GDFP GT YVE YIGRK

c-src    88  ALYDY ESRTET D -------- LA FKK G ERLQIV------MNTEGD WW LAHSLTTGQT GYIP SN YVA PS-DS
v-abl    68  ALYDF VASGDN T -------- LS ITK G EKLRVLG-----YNHNGE WC EAQTK-NGQ- GWVP SN YIT PV-NS
PLC     148  ALFDY KAGRED E -------- LT FTK S AIIQNV------EKQEGG WW RGDYHHKKQ- LWFP SN YVE EMV-S
GAP     284  AILDY TKVPDT D E------- IS FLK G DMFIVN------NELEDG WM WVTNLRTDEQ GLIV ED LVE EV-GR
v-crk   375  ALFDF KGNDDG D -------- LP FKK G DILKIR------DKPEEQ WW NAEDMDGKR- GMIP VP YVE KCRPS
```

FIG. 6

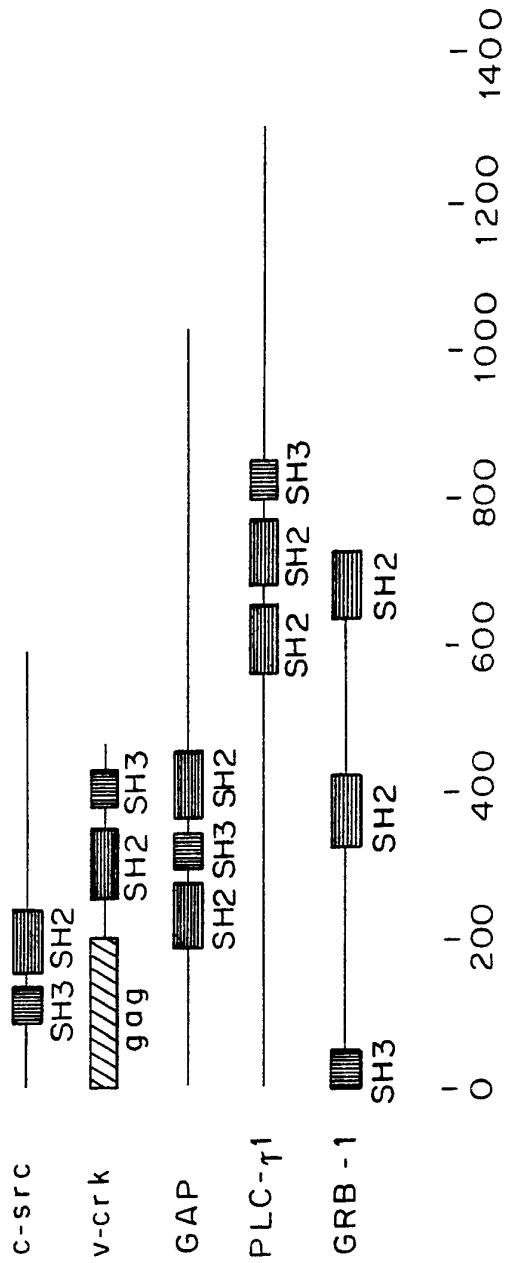
c-src
SH3 SH2
v-crk
gag
SH2 SH3
GAP
SH2 SH3 SH2
PLC-γ1
SH2 SH2 SH3
GRB-1
SH3
SH2
SH2
0
200
400
600
800
1000
1200
1400

FIG. 7

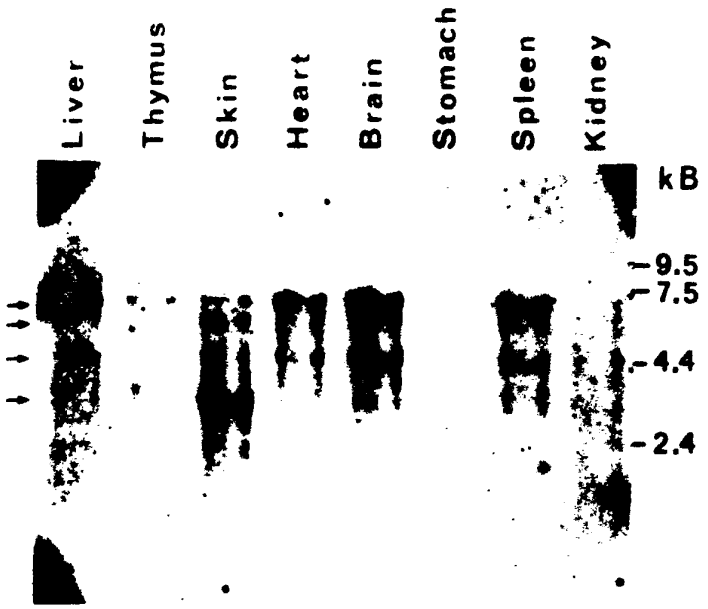
Liver
Thymus
Skin
Heart
Brain
Stomach
Spleen
Kidney
kB
9.5
7.5
4.4
2.4

FIG. 8

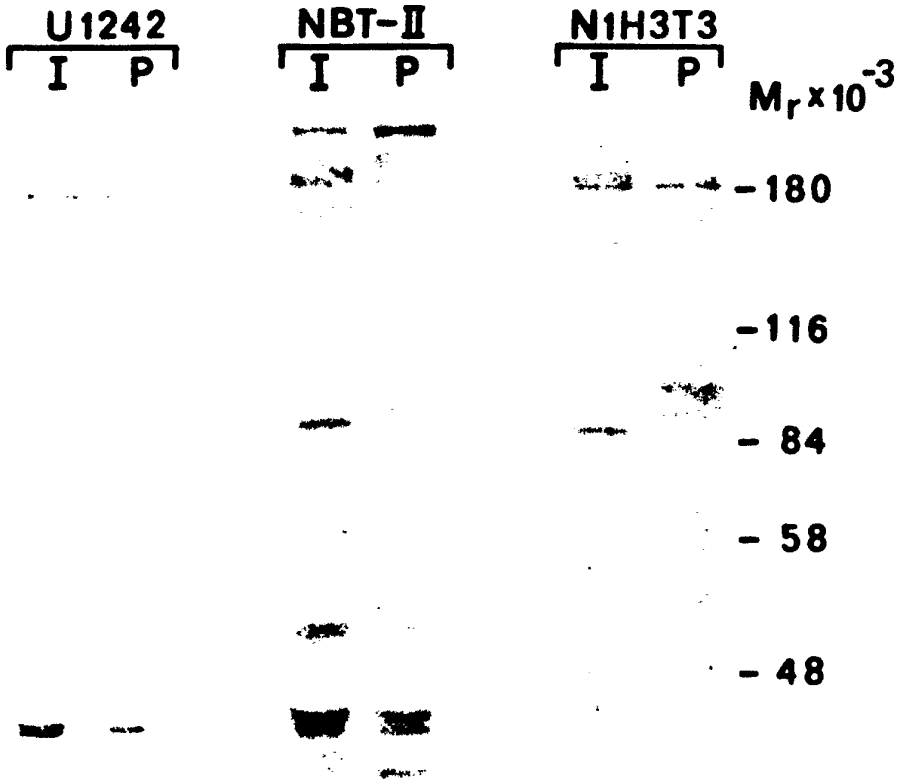
U1242
I P
NBT-II
I P
N1H3T3
I P
Mr x 10-3
180
116
84
58
48

FIG. 9A

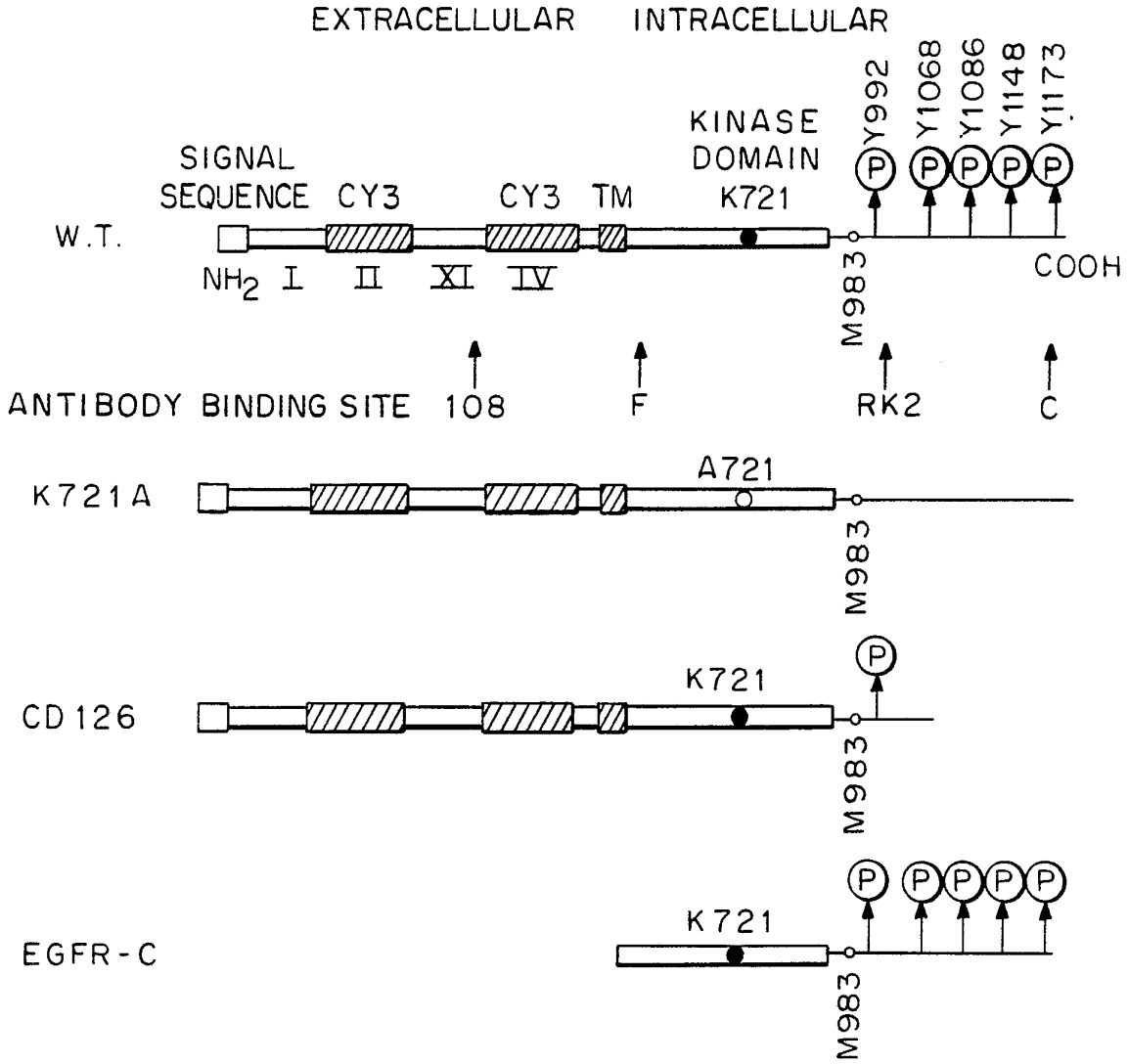
EXTRACELLULAR
INTRACELLULAR
SIGNAL SEQUENCE
CY3
CY3
TM
KINASE DOMAIN
K721
Y992
Y1068
Y1086
Y1148
Y1173
W.T.
NH2
I
II
XI
IV
M983
COOH
ANTIBODY BINDING SITE
108
F
RK2
C
K721A
A721
M983
CD126
K721
M983
EGFR-C
K721
M983

FIG. 9B

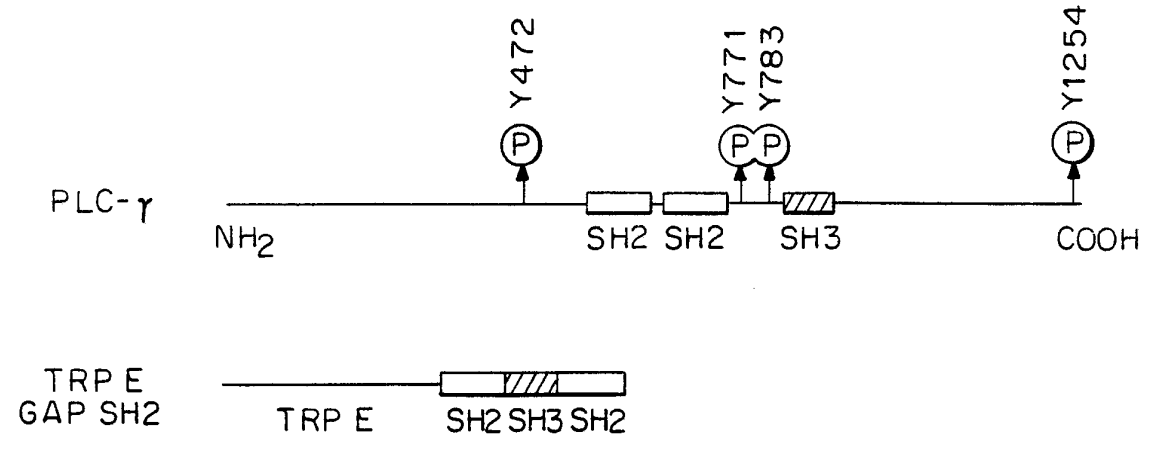
Y472
Y771
Y783
Y1254
PLC-γ
NH2
SH2
SH2
SH3
COOH
TRP E
GAP SH2
TRP E
SH2 SH3 SH2

FIG. 10

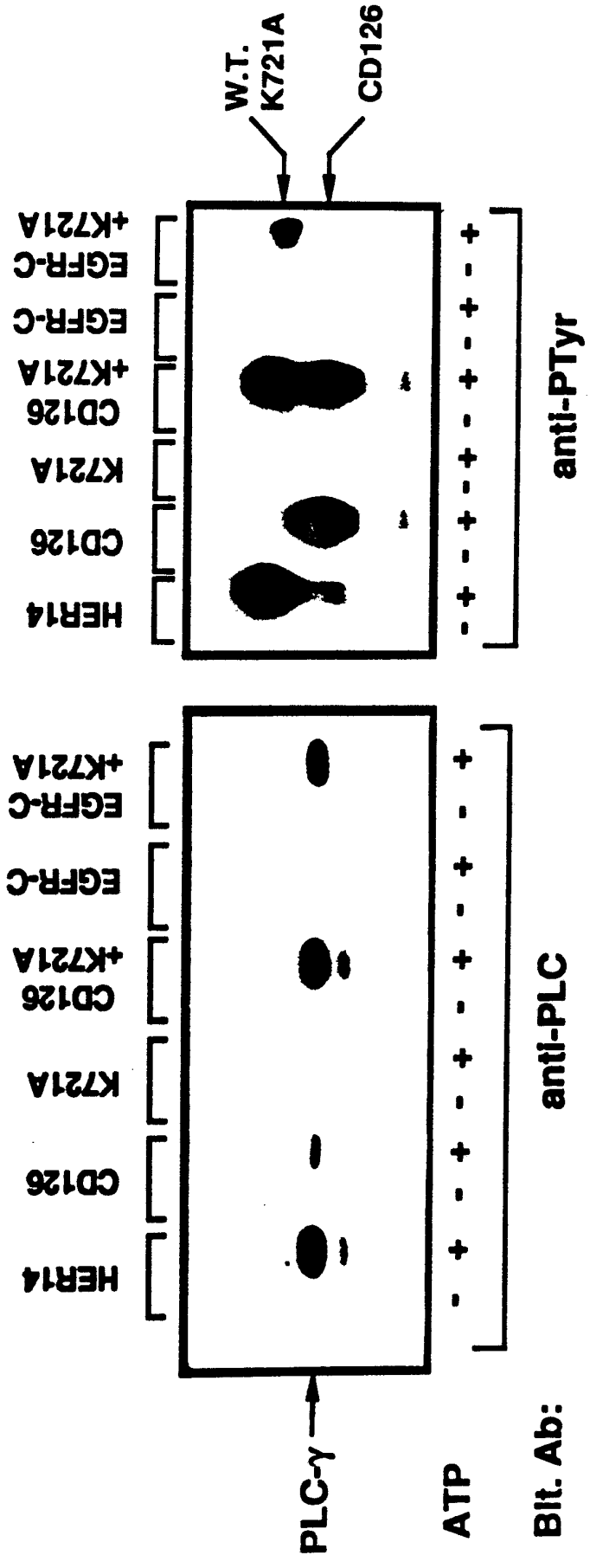
W.T.
K721A
CD126
HER14
CD126
K721A
CD126
+K721A
EGFR-C
EGFR-C
+K721A
anti-PTyr
HER14
CD126
K721A
CD126
+K721A
EGFR-C
EGFR-C
+K721A
anti-PLC
PLC-γ
ATP
Blt. Ab:

FIG. 11

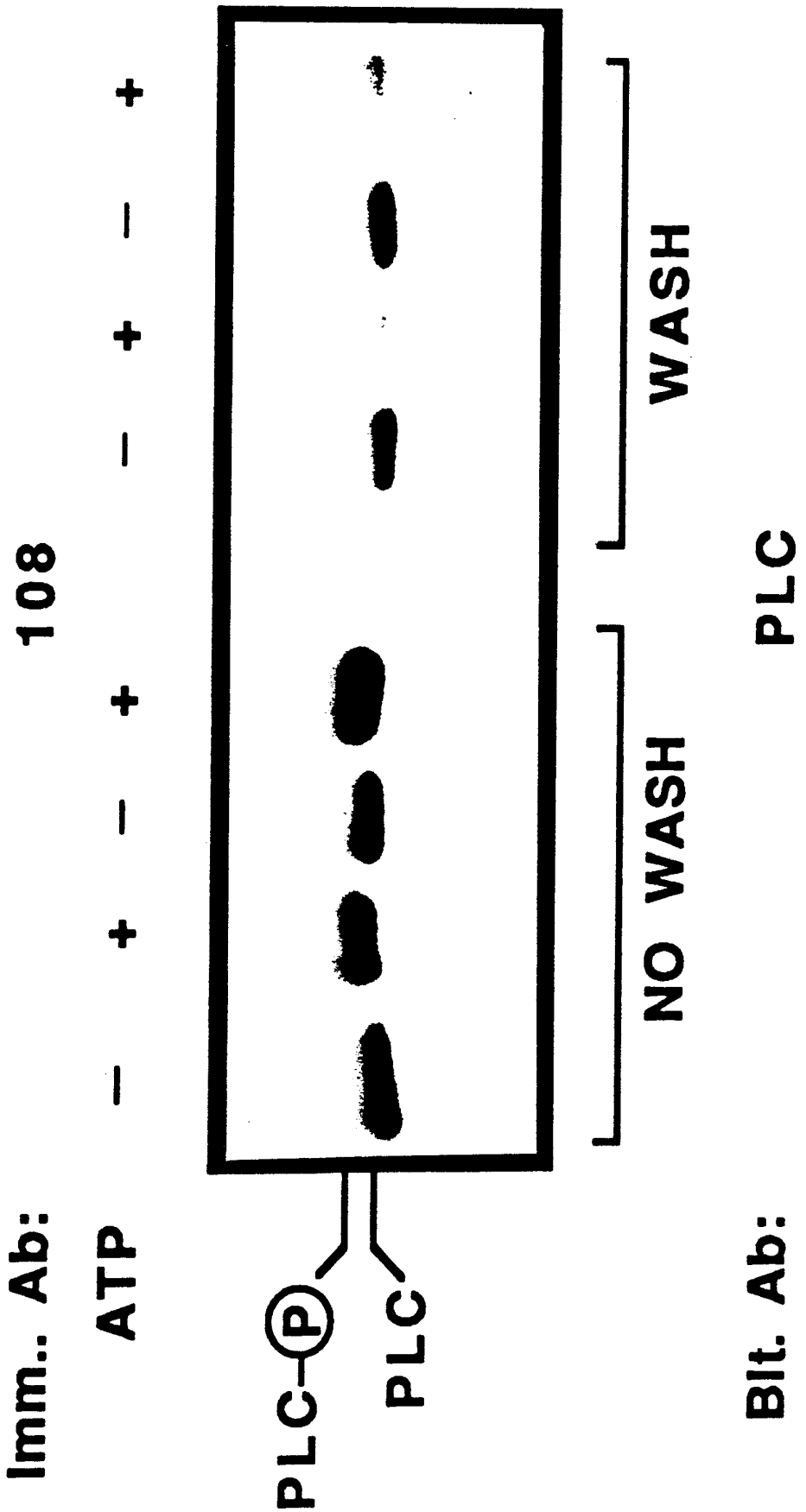

FIG. 12 A

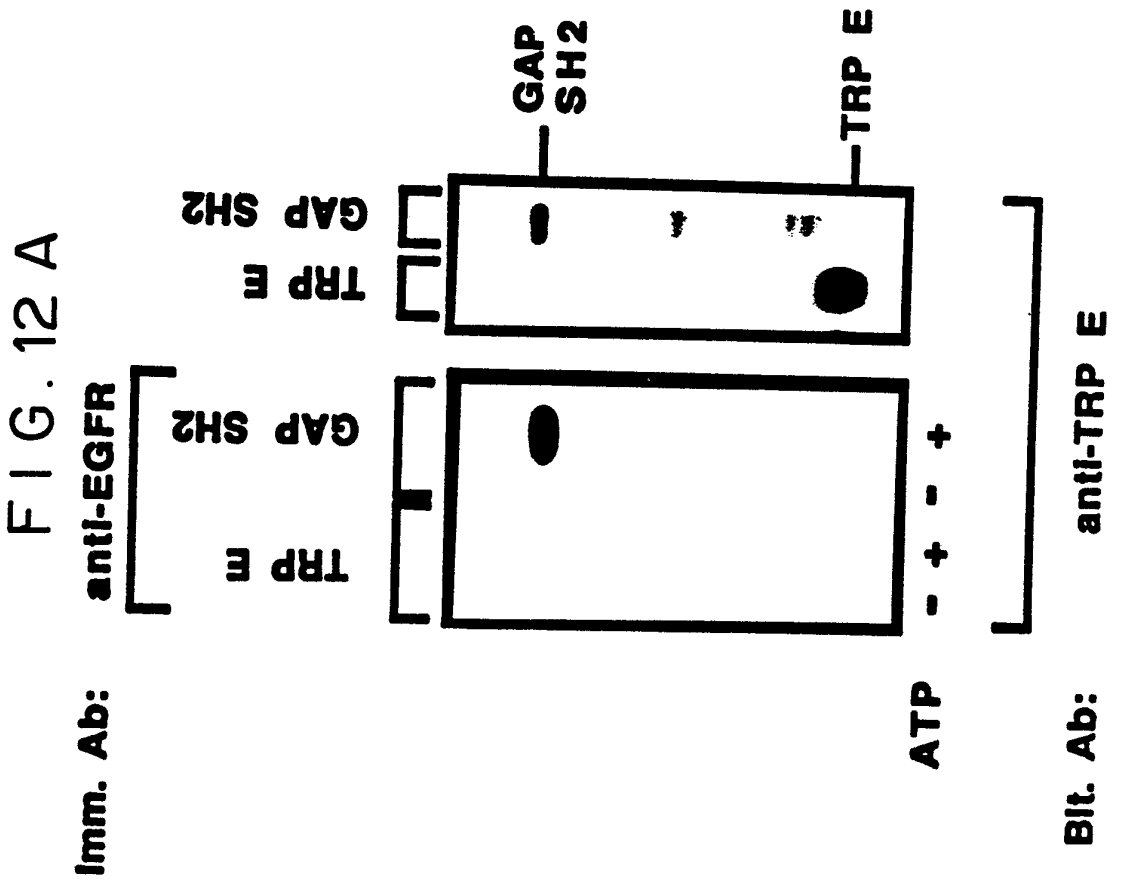

FIG. 12 B

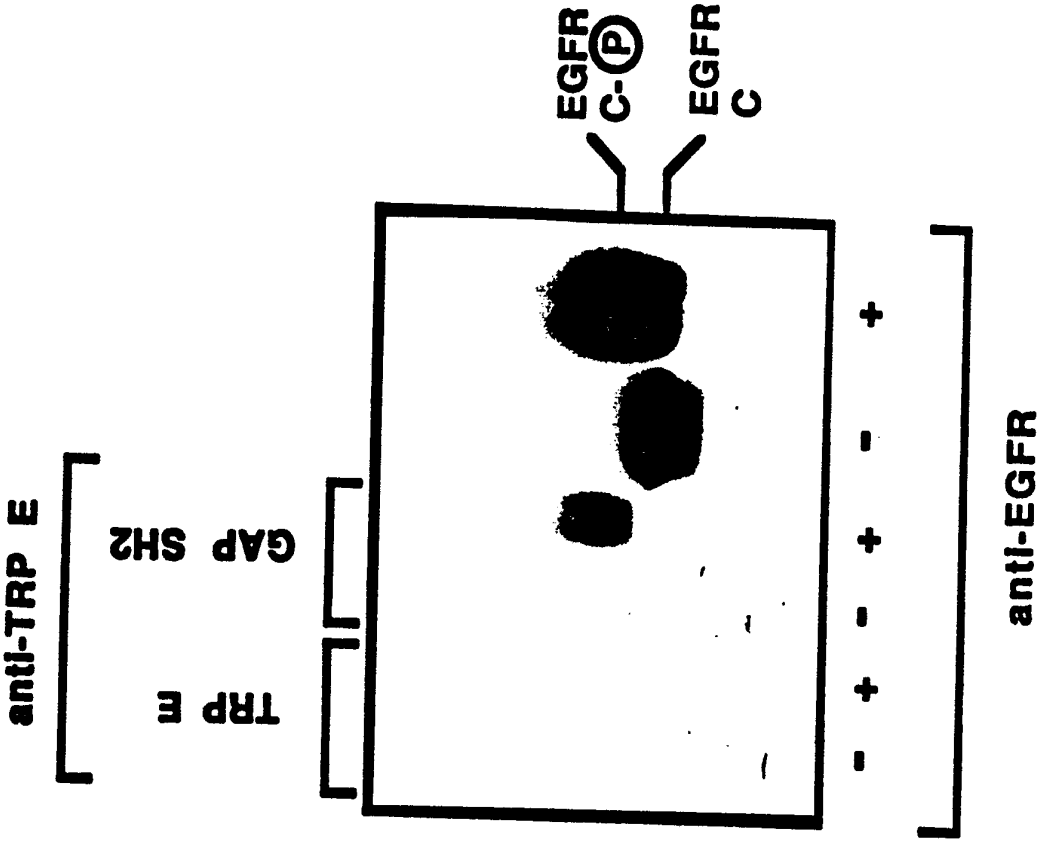

FIG. 13A

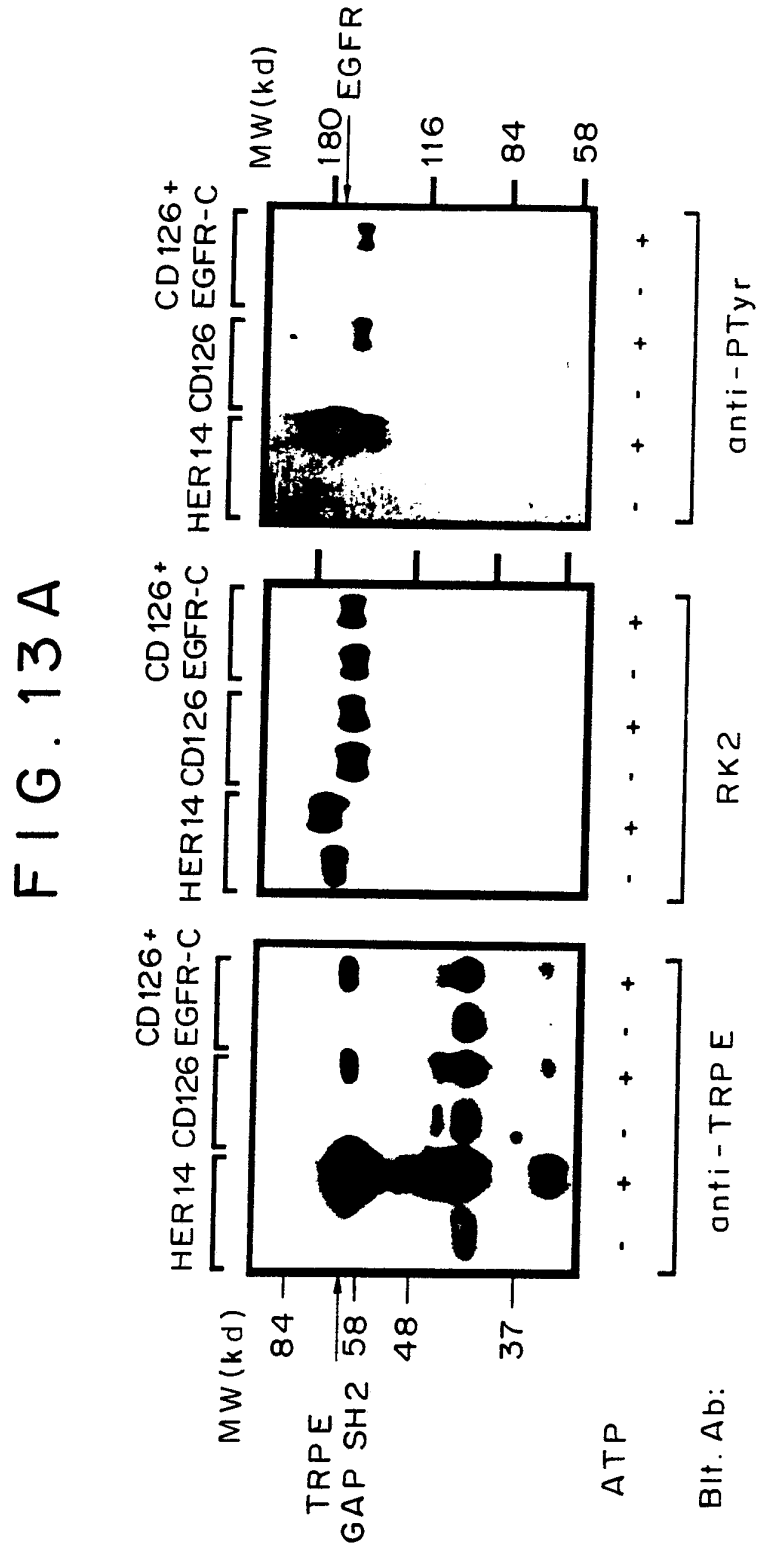
MW(kd)
84
58
48
37
TRPE GAP SH2
ATP
Blt. Ab:
HER14 CD126 EGFR-C
CD126+
anti-TRPE
RK2
anti-PTyr
MW(kd)
180
116
84
58
EGFR

FIG. 13B

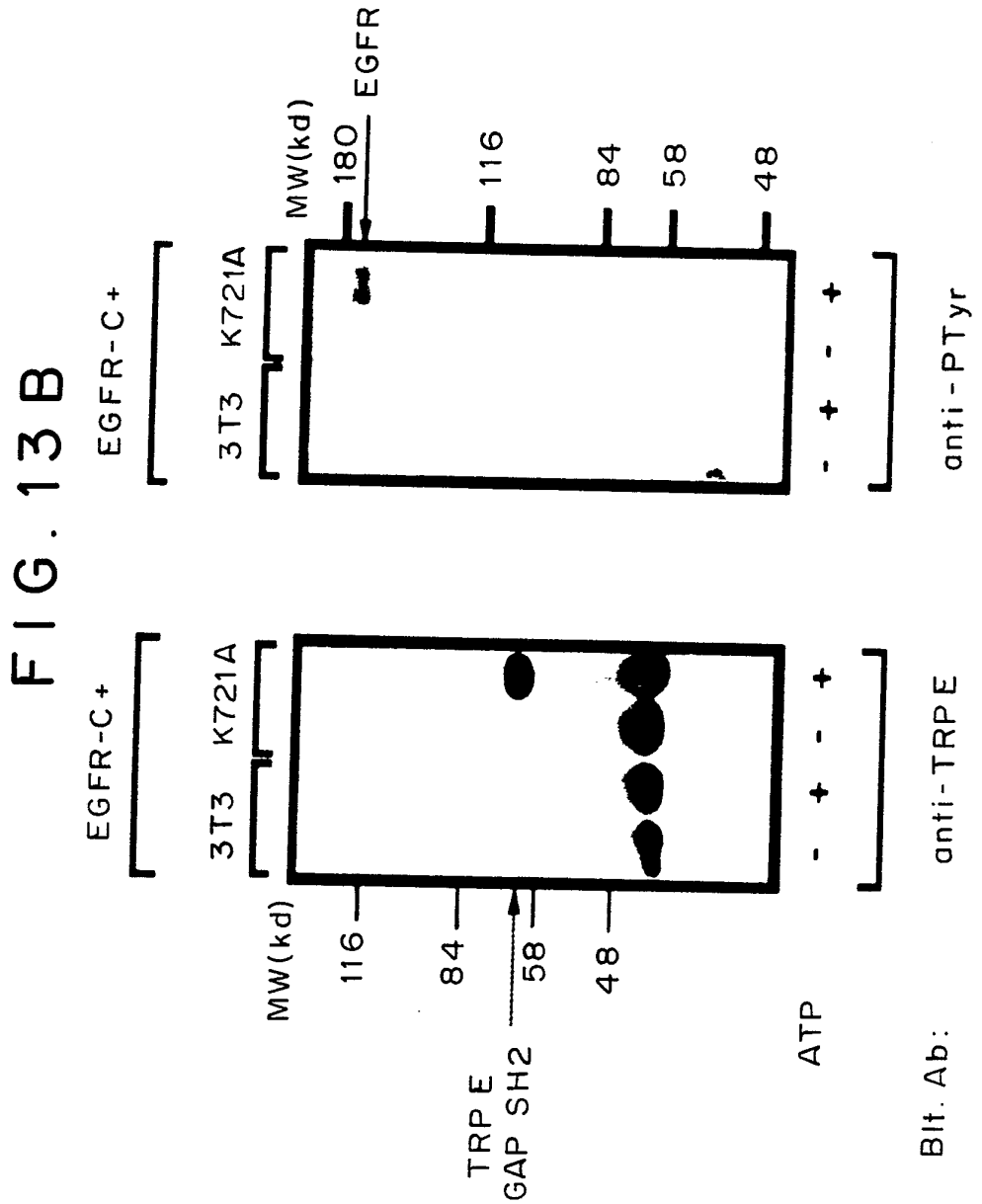
EGFR-C+
3T3 K721A
MW(kd)
180
EGFR
116
84
58
48
- + - +
anti-PTyr
EGFR-C+
3T3 K721A
MW(kd)
116
84
58
48
TRP E
GAP SH2
ATP
- + - +
anti-TRP E
Blt. Ab:

FIG. 14

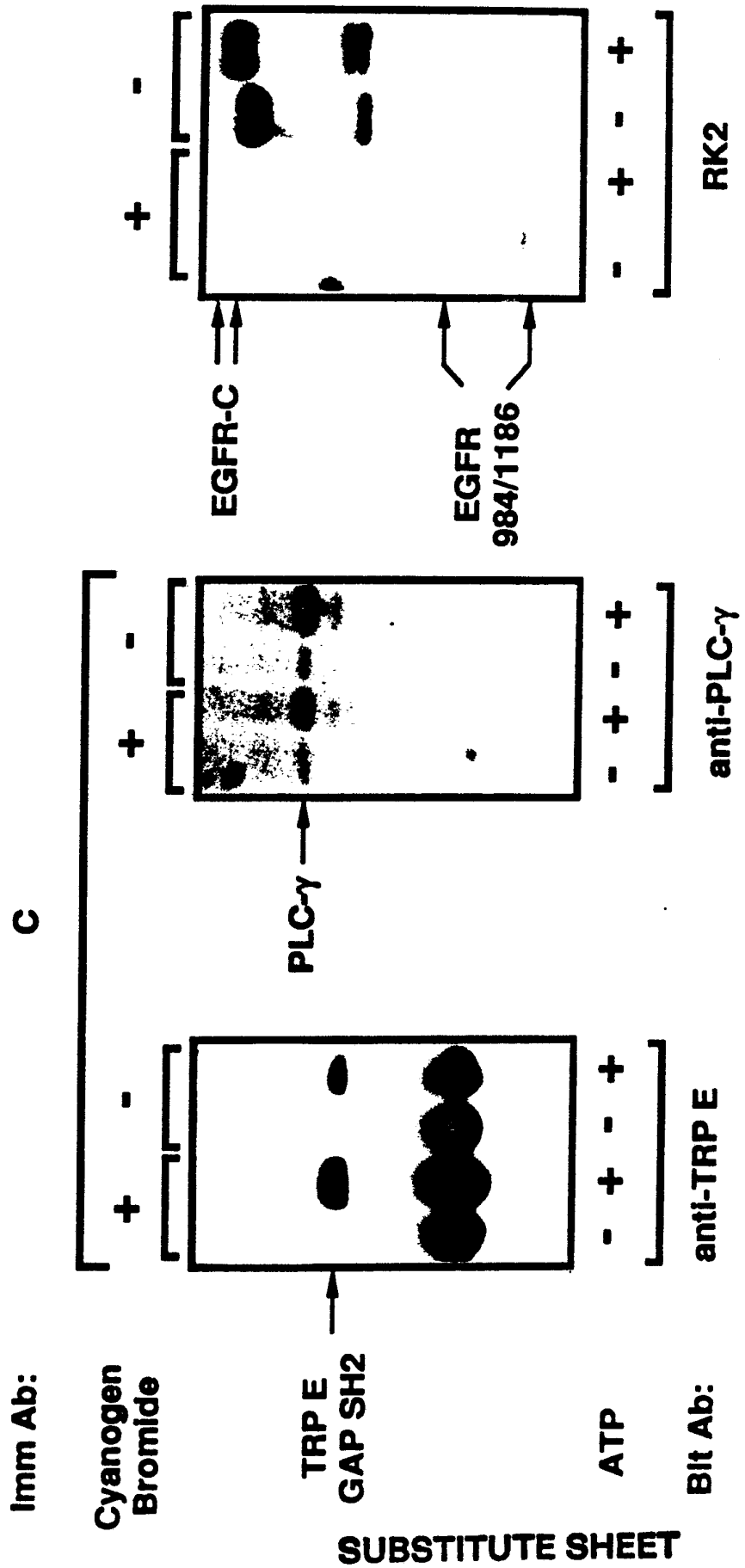
C
Imm Ab:
Cyanogen Bromide
TRP E GAP SH2
ATP
Blt Ab:
anti-TRP E
anti-PLC-γ
PLC-γ
EGFR-C
EGFR 984/1186
RK2

FIG. 15

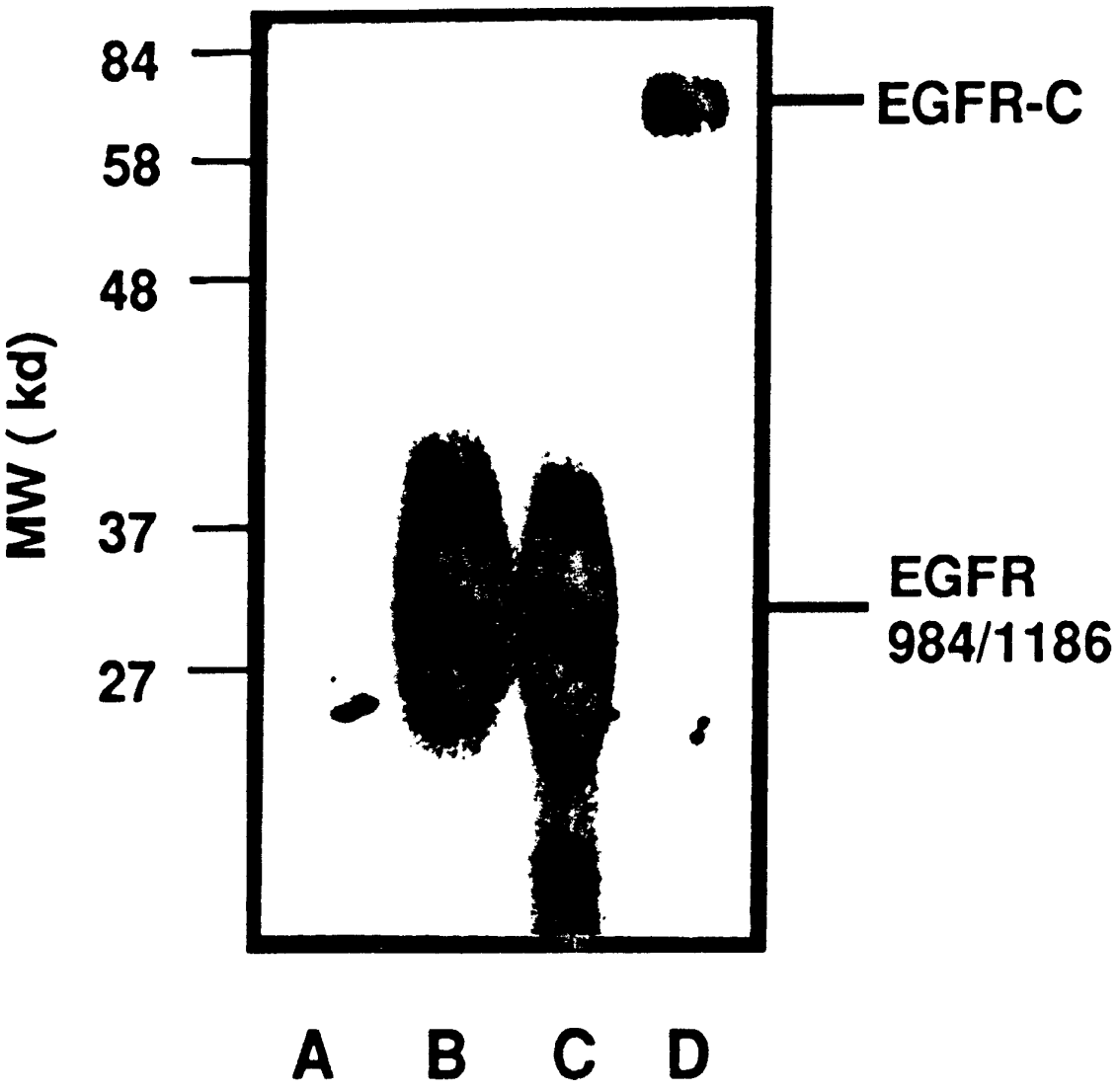

FIG. 16A

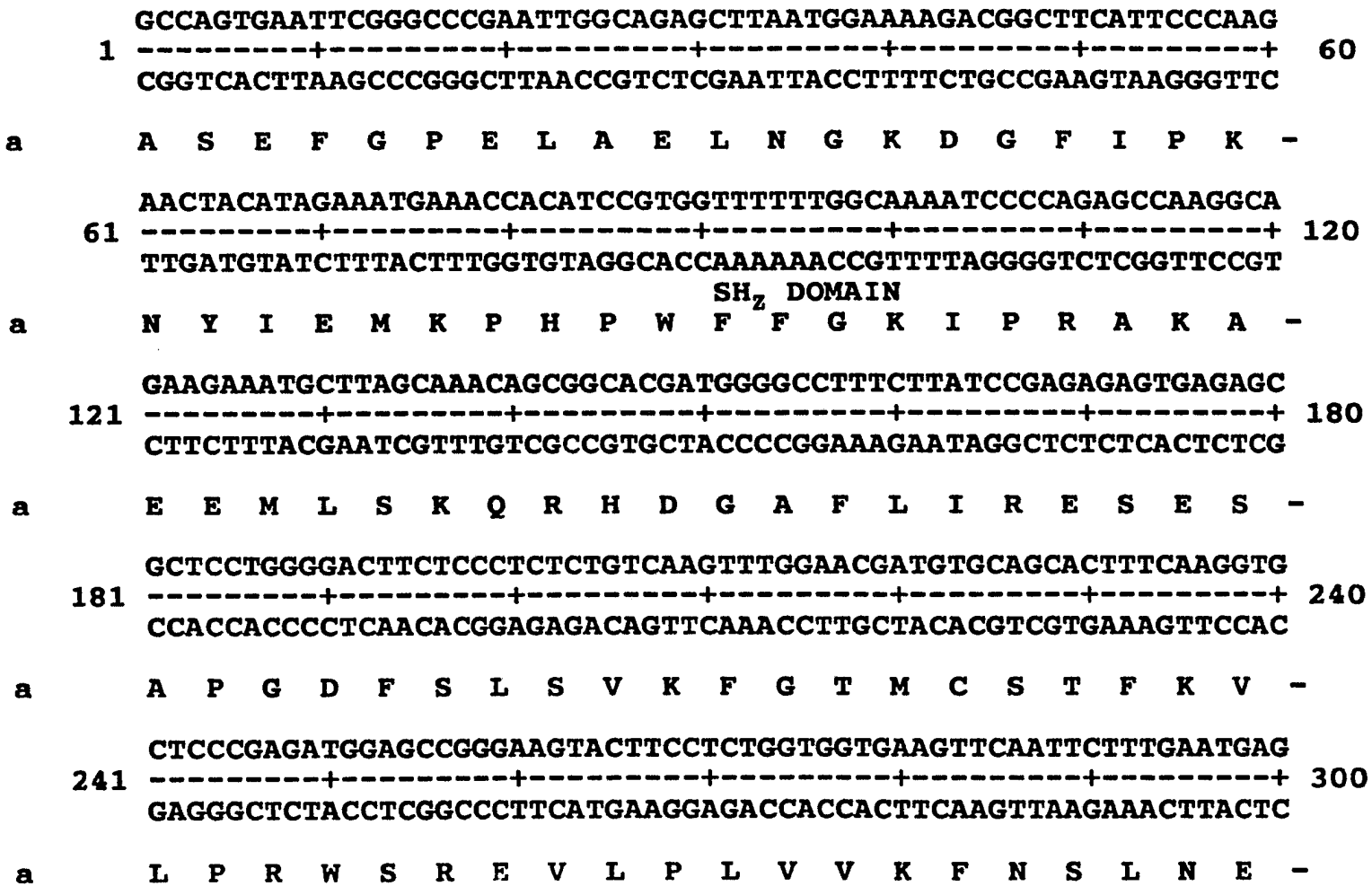
GCCAGTGAATTCGGGCCCGAATTGGCAGAGCTTAATGGAAAAGACGGCTTCATTCCCAAG
1
60
CGGTCACTTAAGCCCGGGCTTAACCGTCTCGAATTACCTTTTCTGCCGAAGTAAGGGTTC
a A S E F G P E L A E L N G K D G F I P K -
AACTACATAGAAATGAAACCACATCCGTGGTTTTTTGGCAAAATCCCCAGAGCCAAGGCA
61
120
TTGATGTATCTTTACTTTGGTGTAGGCACCAAAAAACCGTTTTAGGGGTCTCGGTTCCGT
SH2 DOMAIN
a N Y I E M K P H P W F F G K I P R A K A -
GAAGAAATGCTTAGCAAACAGCGGCACGATGGGGCCTTTCTTATCCGAGAGAGTGAGAGC
121
180
CTTCTTTACGAATCGTTTGTCGCCGTGCTACCCCGGAAAGAATAGGCTCTCTCACTCTCG
a E E M L S K Q R H D G A F L I R E S E S -
GCTCCTGGGGACTTCTCCCTCTCTGTCAAGTTTGGAACGATGTGCAGCACTTTCAAGGTG
181
240
CCACCACCCCTCAACACGGAGAGACAGTTCAAACCTTGCTACACGTCGTGAAAGTTCCAC
a A P G D F S L S V K F G T M C S T F K V -
CTCCCGAGATGGAGCCGGGAAGTACTTCCTCTGGTGGTGAAGTTCAATTCTTTGAATGAG
241
300
GAGGGCTCTACCTCGGCCCTTCATGAAGGAGACCACCACTTCAAGTTAAGAAACTTACTC
a L P R W S R E V L P L V V K F N S L N E -

FIG. 16B

```
     CTGGTGGATTATCACAGATCTACATCTGTCTCCAGAAACCAGCAGATATTCCTGCGGGAC
301  ---------+---------+---------+---------+---------+---------+ 360
     GACCACCTAATAGTGTCTAGATGTAGACAGAGGTCTTTGGTCGTCTATAAGGACGCCCTG

a    L  V  D  Y  H  R  S  T  S  V  S  R  N  Q  Q  I  F  L  R  D  -


     ATAGAACAGGTGCCACAGCAGCCGACATACGTCCAGGCCCTCTTTGACTTTGATCCCCAG
361  ---------+---------+---------+---------+---------+---------+ 420
     TATCTTGTCCACGGTGTCGTCGGCTGTATGCAGGTCCGGGAGAAACTGAAACTAGGGGTC
                                                SH_Z DOMAIN
a    I  E  Q  V  P  Q  Q  P  T  Y  V  Q  A  L  F  D  F  D  P  Q  -

     GAGGATGGAGAGCTGGGCTTCCGCCGGGGAGATTTTATCCATGTCATGGATAACTCAGAC
421  ---------+---------+---------+---------+---------+---------+ 480
     CTCCTACCTCTCGACCCGAAGGCGGCCCCTCTAAAATAGGTACAGTACCTATTGAGTCTG

a    E  D  G  E  L  G  F  R  R  G  D  F  I  H  V  M  D  N  S  D  -

     CCCAACTGGTGGAAAGGAGCTTGCCACGGGCAGACCGGCATGTTTCCCCGCGAATTATGT
481  ---------+---------+---------+---------+---------+---------+ 540
     GGGTTGACCACCTTTCCTCGAACGGTGCCCGTCTGGCCGTACAAAGGGGCGCTTAATACA

a    P  N  W  W  K  G  A  C  H  G  Q  T  G  M  F  P  R  E  L  C  -

     CTCCCCCXGTGAACCGGAACGTCTAAGAGTCAAGAAGCAATTATTTAAAGAAAGTGAAAA
541  ---------+---------+---------+---------+---------+---------+ 600
     GAGGGGGXCACTTGGCCTTGCAGATTCTCAGTTCTTCGTTAATAAATTTCTTTCACTTTT

a    L  P  ?  *  T  G  T  S  K  S  Q  E  A  I  I  *  R  K  *  K  -
```

FIG. 16C

```
    ATGTAAAACACATACAAAAGAATTAAACCCACAAGCTGCCTCTGACAGCAGCCTGTGAGG
601 ---------+---------+---------+---------+---------+---------+ 660
    TACATTTTGTGTATGTTTTCTTAATTTGGGTGTTCGACGGAGACTGTCGTCGGACACTCC

a     M  *  N  T  Y  K  R  I  K  P  T  S  C  L  *  Q  Q  P  V  R  -

    GAGTGCAGAACACCTGGCCGGGTCACCCTGTGACCCTCTCACTTTGGTTGGAACTTTAGG
661 ---------+---------+---------+---------+---------+---------+ 720
    CTCACGTCTTGTGGACCGGCCCAGTGGGACACTGGGAGAGTGAAACCAACCTTGAAATCC

a     E  C  R  T  P  G  R  V  T  L  *  P  S  H  F  G  W  N  F  R  -

    GGGTGGGAGGGGGCGTTGGATTTAAAAATGCCAAAACTTACCTATAAATTAAGAAGAGTT
721 ---------+---------+---------+---------+---------+---------+ 780
    CCCACCCTCCCCCGCAACCTAAATTTTTACGGTTTTGAATGGATATTTAATTCTTCTCAA

a     G  W  E  G  A  L  D  L  K  M  P  K  L  T  Y  K  L  R  R  V  -

    TTTATTACAAATTTTCACTGCTGCTCCTCTTTCCCCTCCTTTGTCTTTTTTTTTTCATCCT
781 ---------+---------+---------+---------+---------+---------+ 840
    AAATAATGTTTAAAAGTGACGACGAGGAGAAAGGGGAGGAAACAGAAAAAAAAAGTAGGA

a     F  I  T  N  F  H  C  C  S  S  F  P  S  F  V  F  F  F  H  P  -

    TTTTTCTCTTCTGTCCATCAGTGCATGACGTTTAAGGCCACGTATAGTCCTAGCTGACGC
841 ---------+---------+---------+---------+---------+---------+ 900
    AAAAAGAGAAGACAGGTAGTCACGTACTGCAAATTCCGGTGCATATCAGGATCGACTGCG

a     F  F  S  S  V  H  Q  C  M  T  F  K  A  T  Y  S  P  S  *  R  -

    CAATAATAAAAACCGAATTCGAGCTCGGATCCGGGGATCCTCTAGAGTC
901 ---------+---------+---------+---------+--------- 949
    GTTATTATTTTTGGCTTAAGCTCGAGCCTAGGCCCCTAGGAGATCTCAG

a     Q  *  *  K  P  N  S  S  S  D  P  G  I  L  *  S  ?  -
```

INTERNATIONAL SEARCH REPORT

International Application No. **PCT/US92/00434**

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[3]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC (5): C07K 13/00, 15/28; C12N 15/62
US CL : 435/69.1, 252.3, 320.1; 530/350, 412

II. FIELDS SEARCHED

Minimum Documentation Searched[4]

Classification System	Classification Symbols
U.S.	435/69.7, 252.3, 320.1; 530/350, 387

Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched[5]

APS, STN/MEDLINE search terms: TYROSINE KINASE#, ASSAY, EXPRESSION CLON?, LAMBDA GT11, ACTIVITY, ENZYMAT?.

III. DOCUMENTS CONSIDERED TO BE RELEVANT[14]

Category*	Citation of Document,[16] with indication, where appropriate, of the relevant passages[17]	Relevant to Claim No.[18]
A	Proceedings of the National Academy of Sciences, Vol.87, issued November 1990, Moran et al,"Src homology region 2 domains direct protein-protein interactions in signal transduction", pages 8622 to 8626, see entire document.	1 to 28
A	METHODS IN ENZYMOLOGY, Vol.146, issued 1987, Pike,"Assay of Growth Factor-Stimulating Tyrosine Kinases Using Synthetic Peptide Substrates", pages 353 to 362, see entire document.	1 to 28

* Special categories of cited documents:[15]

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search[2]	Date of Mailing of this International Search Report[2]
15 APRIL 1992	20 APR 1992
International Searching Authority[1]	**Signature of Authorized Officer**[20]
ISA/US	JOHN D. ULM

Form PCT/ISA/210 (second sheet)(May 1986) ß