Vol. 166, No. 1, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
January 15, 1990 Pages 90-100

EXPRESSION OF EPIDERMAL GROWTH FACTOR RECEPTOR SEQUENCES AS *E. COLI* FUSION PROTEINS: APPLICATIONS IN THE STUDY OF TYROSINE KINASE FUNCTION

John G. Koland, Karen Maddison O'Brien and Richard A. Cerione*

**Department of Pharmacology, NYS College of Veterinary Medicine
Cornell University, Ithaca, New York 14853**

Received November 20, 1989

Summary: To investigate the functions of key domains of the epidermal growth factor receptor (EGFR), various EGFR-derived peptide sequences were expressed in *Escherichia coli* as glutathione S-transferase (GST) fusion proteins. The purified fusion proteins (GST-TK0 - 8) were tested as substrates for the tyrosine kinase activities of the EGFR and c-src. Both the GST-TK4 fusion protein, which contains the major C-terminal tyrosine autophosphorylation sites of the EGFR, and GST-TK7, which contains the connecting sequence between the EGFR kinase domain and the C-terminal autophosphorylation domain, were strongly phosphorylated by the EGFR and c-src. Hence the candidate tyrosine phosphorylation sites present in the connecting sequences of the EGFR, as well as the known autophosphorylation sites of the EGFR, can be phosphorylated by the two tyrosine kinases. The protein GST-TK7 was phosphorylated by c-src with a K_M of 5-10 μM, which indicated a potential interaction between the connecting segment of the EGFR and the c-src kinase. The GST fusion proteins were also used to map the sites recognized by two anti-EGFR monoclonal antibodies and a polyclonal serum raised against an EGFR tyrosine kinase domain fragment. The recognition site of one monoclonal antibody was determined to be in a short sequence surrounding tyr_{1068}, a primary site of autophosphorylation in the C-terminal domain of the receptor. The anti-peptide polyclonal serum recognized only sequences in the GST-TK7 fusion protein, and hence binds to the connecting sequence between the kinase core and the C-terminal domain. These antibodies will therefore be useful reagents for studying the function of two key structural elements of the EGFR tyrosine kinase. The GST-TK fusion proteins should have many other applications in the study of EGFR catalysis and mitogenic signalling.

The epidermal growth factor receptor (EGFR) is in a unique class of receptor proteins which possess an intrinsic protein tyrosine kinase activity (1). The receptor is comprised of a single ~131 kd polypeptide which forms an extracellular growth factor binding domain, an intracellular tyrosine kinase domain (TKD), and a short membrane-spanning sequence which connects the two domains [see Figure 1 and (2,3,4)]. Binding of growth factor to the receptor stimulates the intrinsic protein tyrosine kinase activity and effects a rapid autophosphorylation of specific tyrosine residues of the TKD. The mechanism by which the binding of EGF to the extracellular domain modulates the kinase activity of the intracellular domain is not fully understood. EGF-induced changes in receptor aggregation (2,5,6,7) and in receptor conformation (8,9), the tyrosine autophosphorylation reactions (10), and divalent metal ion activators (8) have each been proposed to have critical roles in the activation process. The conformational changes associated with kinase activation may involve the C-terminal autophosphorylation sites of the TKD. Other structural elements of the TKD are presumably involved in metal ion binding, in substrate recognition, and in the potential interactions of the receptor with transducer proteins.

As as step towards determining how various structural elements of the TKD are involved in receptor activation and signal transduction, we have expressed several peptide sequences of the EGFR TKD as fusion proteins in *Escherichia coli*. Appropriate restriction fragments of the EGFR cDNA sequence were incorporated into pGEX plasmid expression vectors, which include the coding sequences of the *Schistosoma japonicum* glutathione

*Pew Scholar in the Biomedical Sciences, to whom inquiries should be addressed.

Abbreviations: EGF, epidermal growth factor; EGFR, EGF receptor; TKD, tyrosine kinase domain of the EGFR; GST, glutathione S-transferase; GST-TK, family of GST fusion proteins containing EGFR sequences; PMSF, phenylmethylsulfonyl fluoride; IPTG, isopropyl β-D-thiogalactopyranoside; DTT, dithiothreitol; SDS-PAGE, sodium dodecylsulfate polyacrylamide gel electrophoresis.

0006-291X/90 $1.50

S-transferase (GST) gene under control of the *tac* promoter (**11**). Many of the GST fusion proteins were expressed in high yield in *E. coli* and could be readily purified by glutathione-agarose affinity chromatography. We describe the use of the fusion proteins to investigate the substrate specificity of the EGFR and c-src tyrosine kinases, and to map the EGFR TKD sites recognized by polyclonal and monoclonal antibody reagents. These and related GST-TKD fusion proteins should be of significant value in enzymological studies of the EGFR tyrosine kinase, and in determining the means of propagation of mitogenic signals from the receptor.

MATERIALS AND METHODS

Materials- Restriction enzymes, T4 DNA ligase, T4 polynucleotide kinase, and the Klenow fragment of DNA polymerase were purchased from Bethesda Research Laboratories. Lysozyme (grade I), DNaseI (type II), reduced glutathione, glutathione (reduced)-agarose (G-4510), PMSF, and IPTG were obtained from Sigma Chemical Co. [γ^{32}P]ATP was purchased from Dupont-New England Nuclear.

Construction of pGEX-TK bacterial expression vectors- The set of *E. coli* plasmids pGEX-1N, -2T, and -3X were designed for the convenient expression of foreign DNA sequences as glutathione S-transferase (GST) fusion proteins (**11**). Each contains a copy of the *S. japonicum* GST coding sequence (**12,13**) under control of the *tac* promoter. The GST sequences are followed by a multiple cloning site with unique BamHI, SmaI, and EcoRI restriction sites. The multiple cloning sites of the three plasmids are set in three distinct reading frames, so that various cDNA restriction fragments can be ligated in frame. We constructed a family of pGEX-based vectors which expressed various fragments of EGFR cDNA as *E. coli* fusion proteins. The correct construction of the recombinant plasmids was verified by restriction endonuclease mapping, through measurement of the fusion protein molecular weight by SDS-PAGE, and by Western blotting of fusion proteins with anti-EGFR antibodies. The EGFR amino acid sequences coded by the various plasmids, pGEX-TK0 - 8, and the predicted molecular weights of the corresponding fusion proteins, GST-TK0 - 8, are given in Figure 1.

EGFR cDNA fragments were derived from the plasmid pMMTV-EGFR (**14**). The appropriate restriction fragments were isolated by agarose gel electrophoresis and the Geneclean (Bio 101, Inc.) protocol. To construct pGEX-TK0, the 540 bp BamHI fragment of pMMTV-EGFR corresponding to bp 2379-2918 of the EGFR cDNA sequence was cloned into the BamHI site of pGEX-3X. pGEX-TK1 was constructed by ligation of the 769 bp EcoRI fragment of pMMTV-EGFR (EGFR bp 2318-3086) into the EcoRI site of pGEX-1N. The construction of pGEX-TK2 involved the incorporation of C-terminally complete cDNA sequences into pGEX-TK1. The plasmid pGEX-TK1 was cleaved at the unique AatII site of the pGEX vector, and the single-stranded plasmid ends were filled-in with the Klenow fragment of DNA polymerase. The plasmid was cut again at the BglII site of the EGFR cDNA insert, and the large fragment isolated. The 1020 bp fragment of EGFR cDNA generated from the BglII and ScaI digestion of pMMTV-EGFR was isolated and ligated to the large fragment of pGEX-TK1 to yield pGEX-TK2. The plasmid pGEX-TK4 resulted from the deletion of sequences between the EcoRI and NarI sites of pGEX-TK2 (EGFR bp 2318 and 3411 respectively). pGEX-TK2 was first subjected to an incomplete NarI digestion, then a complete EcoRI digestion. The largest fragment, generated by cleavage only in EGFR cDNA sequences, was isolated by agarose gel electrophoresis. The NarI end was filled-in, and the plasmid was recircularized. This fused the remaining C-terminal sequences of EGFR cDNA in frame with the GST sequences of pGEX-1N. The plasmid pGEX-TK6 was constructed by deletion of the EGFR-derived EcoRI fragment (EGFR bp 2318-3086) from pGEX-TK2. The plasmid pGEX-TK7 was derived from pGEX-TK6 by a cleavage at the SstI site (EGFR bp 3294). The SstI ends were filled-in and the plasmid was recircularized. This resulted in the introduction of a termination codon in place of the SstI site. The construction of pGEX-TK8 first required the cloning of the 971 bp BamHI and EcoRI fragment of pMMTV-EGFR (EGFR bp 1348-2318) into the BamHI- and EcoRI-digested pGEX-3X vector. The sequences of the resulting plasmid between the BamHI and NarI sites of the EGFR cDNA (EGFR bp 1348 and 2197 respectively) were removed by incomplete NarI digestion and complete BamHI digestion. The appropriate fragment was gel purified, filled-in and recircularized to yield pGEX-TK8. Blunt-end ligation of the NarI and BamHI ends resulted in the regeneration of the BamHI site and the placement of the EGFR sequences (EGFR bp 2197-2318) in frame with the GST sequences. The construction of pGEX-TK3 required the use of synthetic oligonucleotide adapters so that the 1516 bp ApaI fragment of pMMTV-EGFR (EGFR bp 2146-3661) could be cloned into the BamHI site of pGEX-2T. The BamHI-digested vector was ligated in the presence of a 1000-fold molar excess of 5'-dephosphorylated adapter A (5'-TGACGC-3') and 5'-phosphorylated adapter B (5'-P-GATCGCGTCAGGCC-3'). The adapted vector was isolated by agarose gel electrophoresis, and the isolated ApaI fragment of the EGFR cDNA was cloned into the plasmid. The oligonucleotide adapters maintained the reading frame across the GST-EGFR sequence junction and provided a termination codon at the 3' end of the EGFR sequence.

Purification of GST fusion proteins- Cultures of *E. coli* strain XL1-B (Stratagene) harboring the pGEX-TK plasmids were grown to late log-phase in 750 mL of LB medium supplemented with 100 μg/mL ampicillin at 37°C. Fusion protein expression was induced by the addition of 1 mM IPTG. After an induction period of 15-120 min, the cells were harvested by centrifugation at 4,500 g for 15 min and washed with 100 mL of buffer A (50 mM Tris/HCl, 0.5 mM EDTA, 300 mM NaCl, pH 7.5) supplemented with 0.5 mM PMSF. The cells were resuspended in 10 mL of buffer A plus 5 mM DTT, 1 mM PMSF and 1 mg/mL lysozyme and incubated for 10 min on ice. Five mL of buffer A, supplemented with 5 mM DTT, 1 mM PMSF, 3% Triton X-100, 30 mM $MgCl_2$ and 0.2 mg/mL DNaseI, were then added. After a 10 min incubation on ice, the suspension was centrifuged for 15 min at 60,000 g. The high speed supernatant was recovered and 2 mL of reduced glutathione-agarose in buffer B (20 mM Hepes/Na, 25 mM NaCl, 0.05% Triton X-100, 10% (v/v) glycerol, pH 7.4) were added. The suspension was rocked for 1 h at 3°C, and the agarose gel collected in a 3 mL column. The column was washed with 6 mL of buffer B, and eluted with 6 successive 0.5 mL volumes of buffer B supplemented with 5 mM glutathione (reduced). The fractions with significant protein concentration were pooled. The pool contained predominantly the recombinant fusion protein (0.2-2.0 mg/mL) and was frozen at -70°C until use.

Assays of EGFR and c-src tyrosine kinase activities with fusion protein substrates- Various GST-TK fusion proteins were tested as substrates for the EGFR and c-src tyrosine kinases. The placental EGFR was purified as previously described (**8**), except that a 0.8-0.0 M ammonium sulfate gradient was used in the final elution of the tyrosine agarose affinity column. The recombinant 54 kd form of the EGFR TKD was produced in a recombinant baculovirus/insect cell system (manuscript in preparation). This soluble TKD included EGFR residues 644-1134 and was partially purified by DEAE-

Sephacel chromatography. (Similar results were obtained with an immunoprecipitated preparation of the 54 kd TKD.) The c-src tyrosine kinase was generously provided by Dr. Efraim Racker, Cornell University, Ithaca, NY. This kinase was also expressed in a baculovirus/insect cell system, and was partially purified by HPLC chromatography.

To assay substrate and autophosphorylation reactions, samples of the tyrosine kinase preparations were incubated with the fusion protein substrates for 5 min at room temperature in 10 mM $MgCl_2$ (or 10 mM $MnCl_2$ as indicated), 20 mM Hepes/Na, 25 mM NaCl, 0.05% Triton X-100, 10% (v/v) glycerol, pH 7.4 (31 µL volume) . EGF when added was present at 4 µg/mL. Assays were initiated by the addition of 5 µL of 100 µM [γ^{32}P]ATP (5,000-10,000 dpm/pmol). After the indicated period, the phosphorylation was terminated by the addition of 14 µL of 5xSDS-PAGE sample buffer, and the samples were electrophoresed on a 10% acrylamide/0.13% bisacrylamide SDS gel. The gel was dried, silver stained, and autoradiographed with Kodak AR5 film. In some cases phosphorylation was quantified by the scintillation counting of protein bands excised from the dried gels in Liquiscint solvent (National Diagnostics).

Production of polyclonal rabbit antiserum to a 45 kd recombinant EGFR antigen- The plasmid pET-TK1 was designed for the expression in *E. coli* of a 45 kd TKD fragment of the EGFR (residues 647-1051) (see Figure 7). The 1215 bp NarI fragment of pMMTV-EGFR (EGFR bp 2197-3411) was cloned into the unique NdeI site of the pET-3A expression vector (15), by using the oligonucleotide adapters 5'-CGTTTCA-3' and 5'-P-TATGAAA-3' and the protocol described above for the construction of pGEX-TK3. The resulting plasmid, pET-TK1, was used to transform *E. coli* strain BL21(DE3)/pLysS (15). Although the 45 kd TKD fragment expressed by this system was insoluble and inactive, it was produced in high yield and could be purified by selective solubilization.

A 1 L culture of bacterial cells containing pET-TK1 were grown to log phase, induced for 1 h with 1 mM IPTG, and harvested as described above for the production of GST-TK fusion proteins. The cells were washed in 100 mL of 50 mM Tris/HCl, 1 mM EDTA, 10% (v/v) glycerol, 0.5 mM PMSF, pH 8.0, resuspended in 20 mL of the Tris buffer, and lysed by sonication. The insoluble, recombinant protein was collected by centrifugation at 200,000 g for 1 h, and resuspended in 1 mL of the Tris buffer supplemented with 2% Triton X-100. The suspension was incubated for 15 min on ice, and the insoluble protein again collected by high speed centrifugation. The pellet was solubilized in Tris buffer supplemented with 7 M urea and 5 mM DTT, and contained predominantly the recombinant 45 kd protein (2.5 mg per 1 L culture). An anti-EGFR polyclonal antiserum was raised in rabbit with the recombinant 45 kd TKD fragment of the EGFR as an immunogen. Chinchilla rabbits were immunized subcutaneously with 250 µg of the 45 kd protein in complete Freund's adjuvant, and boosted twice with 100 µg challenges of protein.

Mapping of anti-EGFR antibody recognition sites- The sites recognized by polyclonal and monoclonal antibody preparations were determined by Western blotting of recombinant GST fusion proteins. Crude lysates of *E. coli* cultures carrying the recombinant plasmids, pGEX-TK0 - 4, and the parent plasmid pGEX-1N were prepared. The presence of each recombinant fusion protein, GST-TK0 - 4, and the GST protein was verified by SDS-PAGE and with Coomassie Blue staining. Aliquots (~50 µg) of each lysate were electrophoresed in triplicate on 10% acrylamide/0.27% bisacrylamide SDS gels, transferred to nitrocellulose, and blotted with the rabbit antiserum at a 1/100 dilution or mouse ascites fluid at a 1/500 dilution. The antigen/antibody complexes were detected with protein A-conjugated horse radish peroxidase and a color development in hydrogen peroxide and 4-chloro-1-napthol.

RESULTS

Expression of EGFR sequences in the form of GST fusion proteins- Smith and Johnson (11) have described the use of the plasmid vectors in the pGEX family for the expression of eukaryotic protein sequences as glutathione S-transferase (GST) fusion proteins in *E. coli*. Expression is controlled by the inducible *tac* promoter, and the peptide sequence of interest becomes incorporated downstream of the 27.5 kd GST domain. The fusion proteins tend to be soluble, are produced in high yield, and are rapidly purified by glutathione-agarose affinity chromatography. Some of the pGEX plasmids are engineered so that a protease cleavage site joins the GST protein and the foreign peptide sequence. In these cases, the desired peptide can be freed from the GST domain after purification of the fusion protein.

We have constructed a family of pGEX-based plasmid expression vectors which incorporate specific sequences of the EGFR cDNA into the GST reading frame (see Figure 1). The EGFR cDNA sequences chosen correspond to important peptides in the EGFR TKD (cf. 3,4). For example, the protein GST-TK1 contains the core of the TKD fused to the GST. GST-TK2 includes the full-length C-terminus of the receptor in addition to the tyrosine kinase core. The protein GST-TK4 incorporates an ~15 kd C-terminal sequence of the EGFR, which includes the major sites of tyrosine autophosphorylation (see Discussion). GST-TK7 contains the ~8 kd EGFR sequence which connects the TKD core to the C-terminal domain. The sequence which connects the TKD core to the putative membrane-spanning sequence is found in GST-TK8. This ~5 kd EGFR sequence contains a known site of phosphorylation of the receptor by protein kinase C, thr_{654} (16,17), and also thr_{669}, the site of phosphorylation by a distinct cellular kinase (18,19).

The GST fusion proteins were expressed in high yield, with a 1 L culture yielding 1-10 mg of purified fusion protein. Those proteins containing large portions of the TKD (GST-TK1 - 3) typically were insoluble and inactive as tyrosine kinases (data not shown). The soluble fusion proteins (GST-TK0,TK4 - 8) were effectively purified by glutathione-agarose affinity chromatography (see Materials and Methods). The purification of GST-TK4 as indicated

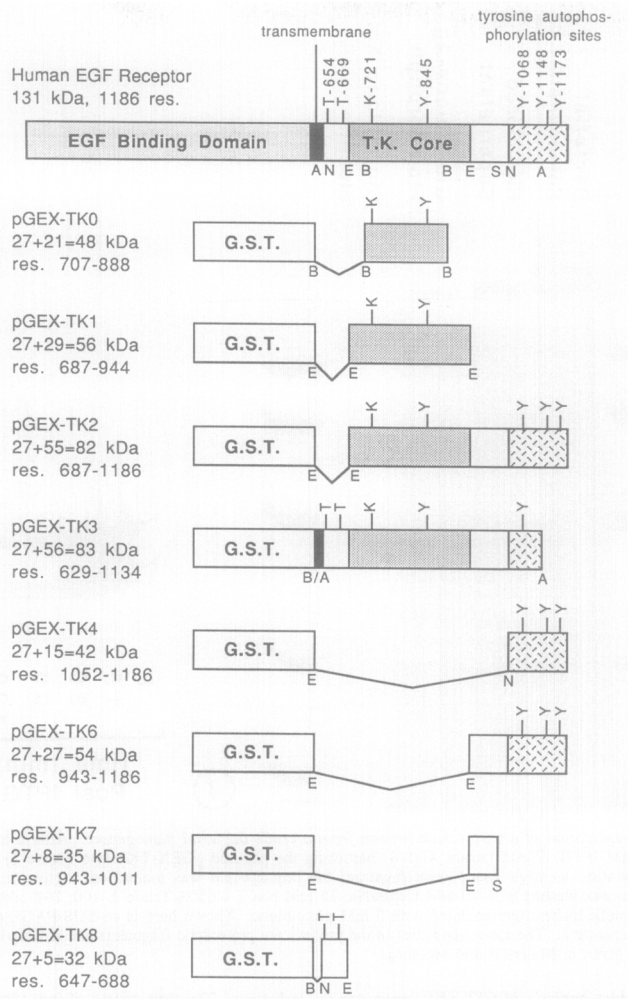

Figure 1. *The schematic structures of the EGFR TKD sequences which have been expressed as GST fusion proteins in E. coli.* Appropriate restriction fragments of EGFR cDNA were isolated and cloned into the pGEX *E. coli* expression vectors (see Materials and Methods). The resulting plasmids pGEX-TK1 - 8 express the indicated EGFR sequences as glutathione S-transferase (GST) fusion proteins under control of the *tac* promoter. The structural elements of the EGFR included in each fusion protein and the fusion protein molecular weights are also indicated. The restriction enzyme sites used in the plasmid constructions are abbreviated as A (ApaI), N (NarI), E (EcoRI), B (BamHI), and S (SstI).

in Figure 2 was typical of the soluble fusion proteins. The insoluble fusion proteins could be purified by selective solubilization of the particulate fraction. Whereas some of the fusion proteins were stable and continued to accumulate throughout the period of induction, certain proteins (GST-TK4 and -TK8) were apparently subject to proteolytic degradation in the bacterium, and their yield was maximized by using shorter periods of induction (see Figure 3). The affinity purification of these proteins yielded some proteolytic fragments in addition to the desired fusion proteins. These fragments contained the intact GST protein, which was apparently resistant to proteolysis. The presence of these fragments or the GST protein itself did not compromise the enzymological and immunological experiments which employed the GST-TK proteins (see below).

GST-TK fusion proteins as substrates for protein tyrosine kinases- Because some of the GST-TK proteins contained known or potential sites of tyrosine autophosphorylation (see Discussion), these proteins were tested as substrates for the EGFR tyrosine kinase (see Figure 4). The protein GST-TK4, which contains the 3 major sites of *in vitro* tyrosine autophosphorylation (20), was strongly phosphorylated by the purified placental EGFR tyrosine kinase. The phosphorylation of GST-TK4 was dependent upon the presence of EGF. The extent of growth factor stimulation was determined to be 3.1-fold, which was very similar to that seen previously with polymeric substrates

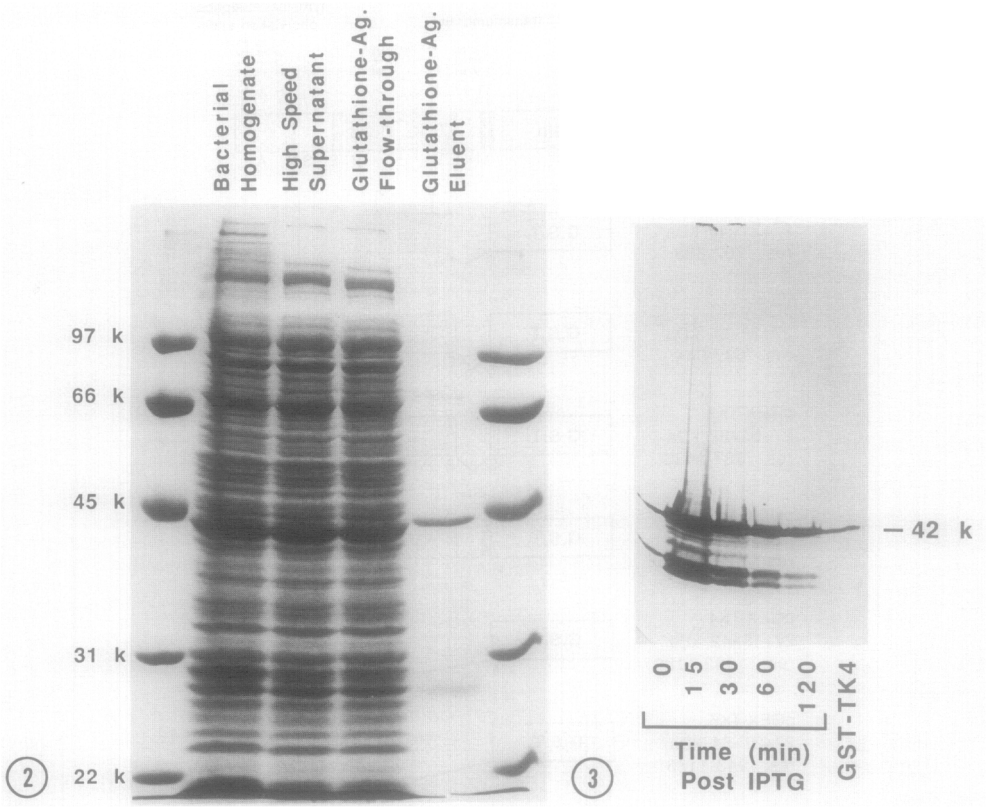

Figure 2. *Typical purification of a GST fusion protein from a crude bacterial homogenate.* After a brief induction of the *tac* promoter by 1 mM IPTG, *E. coli* (strain XL1-B) harboring the plasmid pGEX-TK4 (see Figure 1) were harvested and lysed. After clarification by high speed centrifugation, the homogenate was loaded on a glutathione-agarose affinity column. The column was washed with 20 mM Hepes/Na, 25 mM NaCl, 0.05 % Triton X-100, 10% glycerol, pH 7.4 buffer and eluted with the same buffer supplemented with 5 mM glutathione. Shown here is an SDS-PAGE analysis (Coomassie Blue stain) of the preparation. The main impurities in the product are proteolytic fragments and a dimeric form of the fusion protein. Details are given in Materials and Methods.

Figure 3. *Western blot analysis of GST-TK4 fusion protein induction.* The time course of induction, and the apparent instability of the GST-TK4 fusion protein in the *E. coli* host were analyzed by SDS-PAGE and immunoblotting. After the indicated periods of IPTG induction, *E. coli* XL1-B cultures containing pGEX-TK4 were lysed by the addition of SDS-PAGE sample buffer, and subjected to SDS-PAGE. The proteins were transferred to nitrocellulose, and the GST-TK4 forms were detected by immunoblotting with the mouse monoclonal 291-4A antiserum (see Materials and Methods). The full length GST-TK4 fusion protein and proteolysis products are evident. A sample of the purified GST-TK4 protein was also analyzed.

under these conditions (8). The GST protein itself was not phosphorylated by the EGFR (data not shown), which indicates that the sites of tyrosine phosphorylation of GST-TK proteins must reside on the EGFR-derived sequences. These results are then consistent with the report that the C-terminal sites of the EGFR can be phosphorylated by an intermolecular mechanism (cross-phosphorylation) (21).

The proteins GST-TK6 and -TK7 were also excellent substrates for the EGFR tyrosine kinase (see Figure 5A). GST-TK7 contains the EGFR sequence which connects the C-terminal autophosphorylation domain to the TKD core. Inspection of this sequence indicates the presence of two potential sites of tyrosine autophosphorylation (see Discussion). Apparently at least one of these sites can be strongly phosphorylated by the EGFR, which suggests that the sites in this sequence may become phosphorylated in the intact receptor kinase, either by an intermolecular or autophosphorylation reaction (see Discussion). The fusion protein GST-TK7 was also strongly phosphorylated by a recombinant form of the EGFR, an ~54 kd TKD isolated from a baculovirus/insect cell expression system (see Figure 5B).

The EGFR kinase substrates were also excellent substrates for the c-src tyrosine kinase (see Figure 5A). The c-src kinase did not phosphorylate the GST protein (data not shown), which again indicates that the sites of tyrosine phosphorylation were located on the EGFR-derived sequences of the fusion proteins. The most heavily

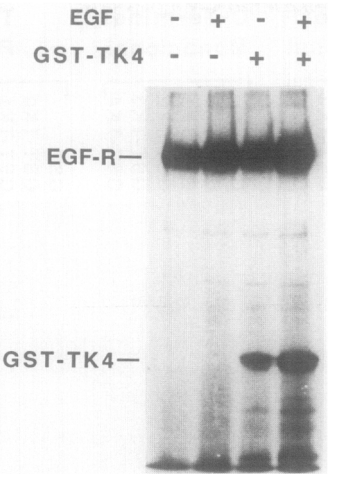

Figure 4. *Phosphorylation of the fusion protein GST-TK4 by the purified EGFR tyrosine kinase.* The fusion protein GST-TK4 (10 μg) (see Figures 1 and 2), which contains the major C-terminal sites of tyrosine autophosphorylation of the EGFR, was incubated with the purified human placental EGFR (0.25 pmol) in the presence of 10 mM $MgCl_2$, and in the presence or absence of excess EGF. The 20 min phosphorylation assays were initiated by the addition of 15 μM [γ^{32}P]ATP, and quenched by the addition of SDS-PAGE sample buffer. Phosphorylation was analyzed by SDS-PAGE, autoradiography, and scintillation counting. Note the autophosphorylation of the EGFR, and that the fusion protein is phosphorylated in an EGF-dependent manner. Details are given in Materials and Methods.

phosphorylated substrate was the GST-TK4 protein, which contains the EGFR C-terminal autophosphorylation sites. This suggests that the EGFR C-terminus may be a substrate for the c-src kinase under some conditions. The EGFR-derived substrate showing the strongest affinity for the c-src kinase was the protein GST-TK7. A K_M of 5-10 μM was estimated for this substrate in the phosphorylation reaction. It is interesting that the mobility of the GST-TK7 after phosphorylation by c-src was consistently lower than after phosphorylation by the EGFR. Since phosphorylation has been known to affect protein mobility in SDS-PAGE (e.g. **22**), it seems likely that the c-src and EGFR tyrosine kinases phosphorylate distinct tyrosine residues in GST-TK7.

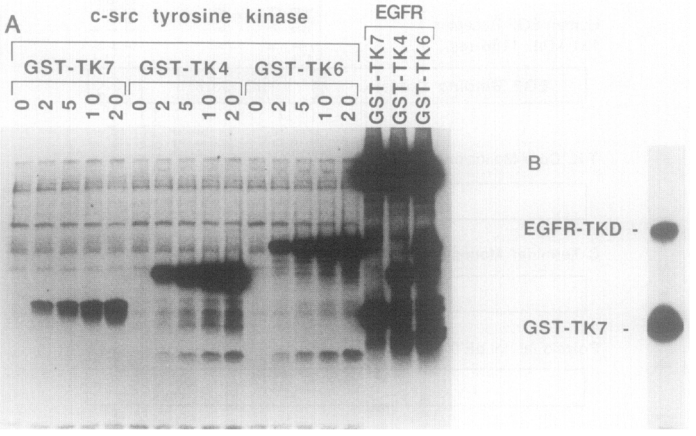

Figure 5. *Phosphorylation of GST-TK fusion protein substrates by the c-src tyrosine kinase, the human placental EGFR, and a 54 kd TKD fragment of the EGFR.* **A.** The phosphorylation of the GST-TK4, -TK6, and -TK7 fusion proteins (see Figure 1) by the c-src and EGFR tyrosine kinases. Baculovirus-expressed c-src (2.5 μg, partially purified) was incubated with the indicated quantities of fusion protein in μg and in the presence of 10 mM $MgCl_2$. The 15 min phosphorylation assays were initiated by the addition of 15 μM [γ^{32}P]ATP, quenched by the addition of SDS-PAGE sample buffer, and analyzed by SDS-PAGE, autoradiography and scintillation counting. Purified placental EGFR (0.25 pmol) was incubated with 20 μg of each fusion protein in the presence of excess EGF and assayed in a similar manner. **B.** Phosphorylation of GST-TK7 by the baculovirus-expressed EGFR TKD. The phosphorylation of the GST-TK7 fusion protein (10 μg) by the 54 kd baculovirus-expressed EGFR TKD (2 μg, partially purified) was assayed in the presence of 10 mM $MnCl_2$ as described in A. Details are given in Materials and Methods.

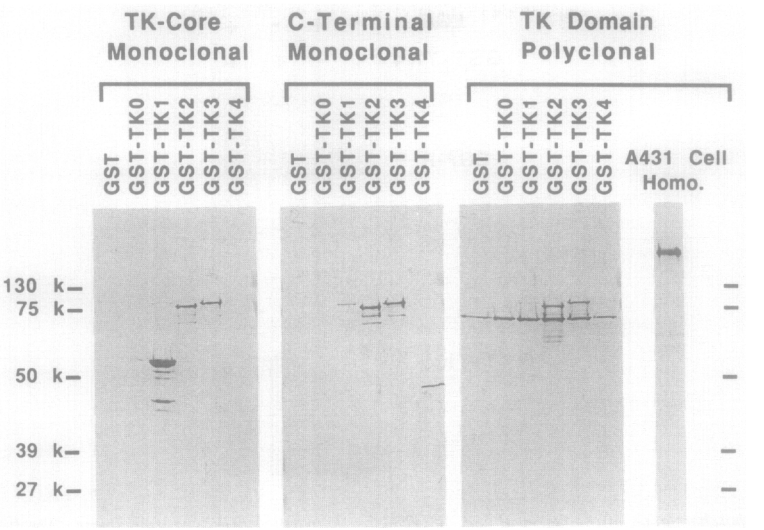

Figure 6. *Interaction of anti-EGFR antibody reagents with the GST-TK fusion proteins.* The sites of interaction of anti-EGFR polyclonal and monoclonal antibodies within the EGFR were determined by immunoblotting of various bacterially expressed GST-TK fusion proteins containing specific portions of the EGFR TKD sequence (see Figure 1). We analyzed two mouse monoclonal antibodies 291-3A and 291-4A, and a rabbit polyclonal IgG fraction generated against a recombinant 45 kd TKD immunogen (see Materials and Methods). A sample of A431 cell plasma membranes, which contain the native 170 kd EGFR was also analyzed. A standard Western blotting protocol with horse radish peroxidase-conjugated second antibody development was used. Note that the rabbit polyclonal also recognized an ~75 kd protein in the *E. coli* lysates.

Mapping of antibody recognition sites with the GST-TK fusion proteins- A useful anti-EGFR rabbit serum was raised with a recombinant 45 kd TKD fragment immunogen (see Materials and Methods). The immunogen was expressed as an insoluble protein in *E. coli* with the plasmid pET-TK1 (see Figure 7). The GST-TK fusion proteins were used to map the sites recognized by this antiserum and two anti-EGFR mouse monoclonal antibodies. Figure 6 shows a typical Western blot analysis used in the mapping of the sequences recognized by the antibodies.

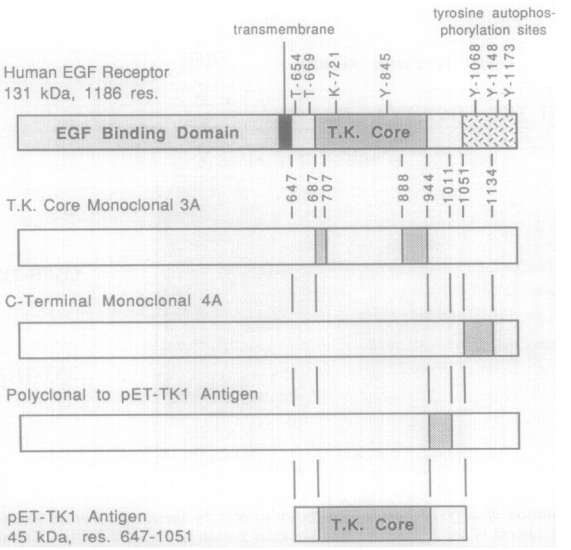

Figure 7. *Map of the sites recognized by polyclonal and monoclonal anti-EGFR antibody reagents.* The results of Figure 6 and similar experiments were used to construct a map of the sites recognized by three antibody reagents. The mouse monoclonal 291-3A reacted with either of the indicated sequences at the ends of the TKD core. The monoclonal 291-4A recognized a short sequence in the C-terminal autophosphorylation domain (res. 1051-1134) which includes tyr_{1068}, a major site of EGFR autophosphorylation. The rabbit polyclonal serum generated against the 45 kd TKD form expressed by plasmid pET-TK1 strongly recognized the "hinge" sequence connecting the C-terminal domain to the TKD core (res. 944-1011). The polyclonal serum did not significantly react with other TKD sequences.

The conclusions of the mapping experiments are summarized in Figure 7. The inner core of the TKD as expressed in GST-TK0 (see Figure 1) was not recognized by any of the antibodies. Neither was the GST protein itself. The rabbit polyclonal serum 8-265 clearly reacted with GST-TK2 and -TK3. (This antiserum also recognized a common ~70 kd contaminant in the *E. coli*-expressed fusion proteins.) In related experiments, the antiserum recognized the proteins GST-TK6 and -TK7, but not GST-TK8 (data not shown). Hence we concluded that the polyclonal serum predominantly reacts with the peptide contained in GST-TK7, the highly acidic sequence connecting the C-terminal sites of autophosphorylation to the TKD and containing residues 943-1011. Figure 6 also demonstrated that the rabbit antiserum can recognize the intact EGFR in a background of mammalian cell proteins, namely the plasma membrane proteins of A431 carcinoma cells.

We were also able to delimit the sites of the EGFR recognized by two useful monoclonal antibodies, 291-3A and 291-4A produced and provided by Dr. R. Schatzman (Syntex Research). Because the monoclonal 291-4A recognized the fusion proteins GST-TK3 and -TK4, the site of recognition by this antibody must include residues 1052-1134, a short C-terminal sequence which includes tyr_{1068}, one of the three major sites of *in vitro* autophosphorylation (see Discussion). Because the other monoclonal 291-3A recognized GST-TK1, but not GST-TK0, we could limit its sites of recognition to two short sequences between residues 687-706 and 889-944 which are located at the two ends the TKD core.

DISCUSSION

Previous enzymological investigations of the EGFR have highlighted the roles of specific amino acid residues in receptor catalysis and regulation [the cDNA and amino acid sequences of the EGFR are given in (4)]. The core of the TKD contains lys_{721} (see Figure 1), which is presumably located within the nucleotide binding site and is essential for tyrosine kinase activity (23,24). Of much interest have been the roles of the three known sites of *in vitro* tyrosine autophosphorylation, tyr_{1068}, tyr_{1148} and tyr_{1173} (20), which are located in an ~13 kd C-terminal sequence of the receptor. Other potential sites of tyrosine phosphorylation are evident in the EGFR sequence (see Table I). These sites are indicated by the presence of a negatively charged amino acid(s) 1 or 4 residues N-terminal to the tyrosine and include tyr_{845}, tyr_{974}, tyr_{992}, and tyr_{1114}. tyr_{845} is absolutely conserved among tyrosine kinases of known sequence (25), and the corresponding residue is phosphorylated in the activated src kinase (26). The major site of *in vivo* tyrosine autophosphorylation of the EGFR has been determined to be tyr_{1173} (20). Given that a particular tyrosine phosphorylation may be transient and/or labile due to the presence of cellular phosphatases, it seems plausible that some of the other potential phosphorylation sites may be involved in EGFR function. Indeed when phenylalanine was substituted for tyr_{1173} in a recombinant EGFR mutant, a strong *in vivo* autophosphorylation of the receptor was maintained (27). Recent work has indicated that the c-src tyrosine kinase is phosphorylated on a C-terminal site by a distinct but unknown cellular tyrosine kinase (28). This leads to the speculation that the multiple tyrosine phosphorylation sites of the EGFR C-terminus may be also be phosphorylated by another tyrosine kinase, for example by the c-src kinase.

As a convenient means of addressing these issues concerning the phosphorylations of the EGFR, and to facilitate immunological approaches to tyrosine kinase structure and function (see below), we developed the GST-TK family of plasmid expression vectors. These *E. coli* vectors enabled the purification of EGFR-derived peptides in the form of GST fusion proteins (11). The peptides selected for expression included important structural and functional domains of the EGFR. Those fusion proteins containing short (<25 kd) sequences of the EGFR tended to be soluble, were relatively stable, and could be purified by affinity chromatography. Fusion proteins containing larger sequences of the EGFR were found in the particulate fraction. These proteins could be purified by selective solubilization, and although inactive, were useful for immunological mapping of determinant sites.

Our preliminary phosphorylation studies showed that those GST-TK proteins which contain candidate autophosphorylation sites of the EGFR TKD were useful substrates for the EGFR tyrosine kinase and the related c-src tyrosine kinase. The protein GST-TK4, which contains the C-terminal autophosphorylation domain of the EGFR and its four candidate autophosphorylation sites, was strongly phosphorylated by both the EGFR and c-src

Table I

Sequences of Potential Sites of Tyrosine Phosphorylation in the EGFR

Sequence of Site[a]	EGFR Location	Relevant GST Fusion Proteins	Notes
glu-glu-lys-glu-tyr$_{845}$	TKD core	GST-TK0,1	highly conserved among tyrosine kinases
asp-ser-asn-phe-tyr$_{974}$	"hinge"[b]	GST-TK6,7	
asp-ala-asp-glu-tyr$_{992}$	"hinge"	GST-TK6,7	
pro-val-pro-glu-tyr$_{1068}$	C-term	GST-TK3,4	P3[c], known site *in vitro*
gly-asn-pro-glu-tyr$_{1114}$	C-term	GST-TK3,4	
asp-asn-pro-asp-tyr$_{1148}$	C-term	GST-TK4	P2, known site *in vitro*
glu-asn-ala-glu-tyr$_{1173}$	C-term	GST-TK4	P1, known site *in vitro* and *in vivo*

[a]The amino acid sequence and numbering conventions are those given by (4).
[b]The sequence of ~8 kd connecting the C-terminal autophosphorylation domain to the TKD and comprising the EGFR sequence in GST-TK7 are informally referred to as the "hinge."
[c]P1, P2, and P3 are the designations for known sites of *in vitro* autophosphorylation (20).

tyrosine kinases. The phosphorylation of this domain by c-src suggests a potential cross-phosphorylation of the two kinases under physiological conditions. The strong phosphorylation of the protein GST-TK7 by the EGFR tyrosine kinase was also intriguing. This fusion protein contains the ~8 kd acidic EGFR sequence which connects the C-terminal autophosphorylation domain to the TKD core. Whereas the amino acid sequence of this peptide indicates the presence of potential tyrosine autophosphorylation sites (see Table I), no phosphorylation of tyrosines in this "hinge" sequence has been detected in studies of the intact EGFR (20), although the possibility of such phosphorylation has not been excluded. If this sequence indeed functions in the intact EGFR as a hinge, such that it is involved in structural displacements of the autophosphorylation domain, then the autophosphorylation of tyrosine residues in this sequence could regulate the further autophosphorylation or substrate phosphorylation reactions of the tyrosine kinase.

The protein GST-TK7 was also phosphorylated by the c-src tyrosine kinase. In contrast to the substrate GST-TK4, the GST-TK7 fusion protein showed a strong affinity for the c-src kinase, with a K_M of 5-10 μM being indicated. This K_M is among the lowest reported for substrates of the c-src kinase (25). It is possible that the presence of a highly acidic sequence in the EGFR-derived portion of the GST-TK7 substrate (i.e. the sequence from residues 979 to 991 in which 9 of 13 residues are glutamate or aspartate) is responsible for the high affinity of the substrate. A homologous acidic sequence in the middle T antigen has been reported to be responsible for the interaction of the tumor antigen with c-src (29,30). Furthermore, the src-associated middle T antigen is phosphorylated on a tyrosine residue located near the acidic sequence (31). The interaction of GST-TK7 with c-src may also be related to the observations that c-src expression can potentiate the mitogenic effects of EGF in cultured cells (32).

Several polyclonal and monoclonal antibodies are available for the study of the EGFR function. We have used the GST-TK fusion proteins to map the sites of interaction of two anti-EGFR monoclonal antibodies and a polyclonal rabbit antiserum generated with a bacterially expressed 45 kd TKD fragment as the immunogen (see Figure 7). The sites of interaction of these antibodies with the EGFR include important structural elements of the EGFR cytoplasmic domain. The polyclonal serum reacted strongly with the "hinge" sequence of the TKD (residues 944-1011) which was expressed in the GST-TK7 fusion protein. This is reasonable given the highly acidic nature of

this sequence and the correlation of charge with immunogenicity. Other sequences of the EGFR kinase domain showed little reactivity with the polyclonal serum. Although the sites of interaction of the polyclonal antiserum under the denaturing conditions of Western blotting may differ from those recognized in the native TKD structure, the polyclonal serum may prove to be useful for studying the role of the "hinge" region in TKD regulation or in the interaction of the TKD with cellular substrates.

The sites of interaction of the 291-3A and 291-4A mouse monoclonal antibodies were also mapped (see Figure 7). The 291-3A antibody was determined to recognize either one of the two short sequences at the ends of the TKD core. The site of interaction of the monoclonal antibody 291-4A was determined to be within the sequence from residues 1051 to 1134 in the C-terminal domain of the EGFR. This sequence contains one of the three major sites of *in vitro* tyrosine autophosphorylation, tyr_{1068}. Hence this latter C-terminal-specific monoclonal may be useful for studying the role of autophosphorylation in EGFR-mediated events.

In conclusion, the expression of EGFR-derived peptide sequences as GST fusion proteins in *E. coli* has been shown to be a useful approach for the study of the roles of these sequences in tyrosine kinase structure and function. The GST-TK fusion proteins were expressed in relatively high yield and could be readily purified. We have shown that, in contrast to the parent GST protein, the GST-TK fusion proteins carrying the predicted autophosphorylation loci of the EGFR (see Table I) were strongly phosphorylated by the EGFR and c-src tyrosine kinases. Hence these and related substrates may prove to be of much value in the study of tyrosine kinase catalysis. The possibility of generating by site-specific mutagenesis substrates with specific amino acid alterations is attractive. The fusion proteins were also found to be useful for the mapping of anti-EGFR antibody recognition sites. The mapping of these sites will greatly increase the utility of these antibody reagents in determining the roles of specific structural elements of the EGFR cytoplasmic domain. In principle the fusion proteins would also be useful for the generation of other site-specific antibodies to the EGFR. Finally, investigators of other protein kinases might also consider the expression of the key sequence elements of the kinase or kinase substrates as GST fusion proteins.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institutes of Health (GM40654), the Pew Biomedical Research Scholars Program and the Cornell Biotechnology Program, which is supported by the New York State Science and Technology Foundation, a consortium of industries, and the United States Army. We thank Dr. Randy Schatzman for providing the monclonal antibodies used in this work.

REFERENCES

1. Yarden, Y., and Ullrich, A. (1988) *Ann. Rev. Biochem.* **57**, 443-478.
2. Schlessinger, J. (1986) *J. Cell Biol.* **103**, 2067-2072.
3. Hunter, T. (1984) *Nature* **311**, 414-416.
4. Ullrich, A., Coussens, L., Hayflick, J. S., Dull, T. J., Gray, A., Tam, A. W., Lee, J., Yarden, Y., Libermann, T. A., Schlessinger, J., Downward, J., Mayes, E. L. V., Whittle, N., Waterfield, M. D., and Seeburg, P. H. (1984) *Nature* **309**, 418-425.
5. Biswas, R., Basu, M., Sen-Majumdar, A., and Das, M. (1985) *Biochemistry* **24**, 3795-3802.
6. Yarden, Y., and Schlessinger, J. (1987) *Biochemistry* **26**, 1434-1442.
7. Boni-Schnetzler, M., and Pilch, P. F. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 7832-7836.
8. Koland, J. G., and Cerione, R. A. (1988) *J. Biol. Chem.* **263**, 2230-2237.
9. Staros, J. V., Cohen, S., and Russo, M. W. (1985) in *Molecular Mechanisms of Transmembrane Signalling* (Cohen, P. and Houslay, M. D., eds) pp. 253-278, Elsevier.
10. Bertics, P. J., and Gill, G. N. (1985) *J. Biol. Chem.* **260**, 14642-14647.
11. Smith, D. B., and Johnson, K. S. (1988) *Gene*, **67**, 31-40.
12. Smith, D. B., Davern, K. M., Board, P. G., Tiu, W. U., Garcia, E. G., and Mitchell, G. F. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8703-8707.
13. Smith, D. B., Davern, K. M., Board, P. G., Tiu, W. U., Garcia, E. G., and Mitchell, G. F. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6541.
14. Clark, A. J. L., Beguinot, L., Ishii, S., Ma, D. P., Roe, B. A., Merlino, G. T., and Pastan, I. (1986) *Biochim. Biophys. Acta* **867**, 244-251.
15. Rosenberg, A. H., Lade, B. N., Chui, D., Lin, S.-W., Dunn, J. J., and Studier, F. W. (1987) *Gene* **56**, 125-135.
16. Hunter, T., Ling, N., and Cooper, J. A. (1984) *Nature* **311**, 480-483.
17. Davis, R. J., and Czech, M. P. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1974-1978.
18. Heisermann, G. J., and Gill, G. N. (1988) *J. Biol. Chem.* **263**, 13152-13158.
19. Countaway, J. L., Northwood, I. C., and Davis, R. J. (1989) *J. Biol. Chem.* **264**, 10828-10835.

20. Downward, P. P., and Waterfield, M. D. (1984) *Nature* **311**, 483-485.
21. Honegger, A. M., Kris, R. M., Ullrich, A., and Schlessinger, J. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 925-929.
22. Yonemoto, W., Jarvis-Morar, M., Brugge, J. S., Bolen, J. B., and Israel, M. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4568-4572.
23. Honegger, A. M., Dull, T. J., Felder, S., Van Obberghen, E., Bellot, F., Szapary, D., Schmidt, A., Ullrich, A., and Schlessinger, J. (1987) *Cell* **51**, 199-209.
24. Chen, W. S., Lazar, C. S., Poenie, M., Tsien, R. Y., Gill, G. N., and Rosenfeld, M. G. (1987) *Nature* **328**, 820-823.
25. Hunter, T., and Cooper, J. A. (1986) *The Enzymes* **17**, 191-246.
26. Smart, J. E., Oppermann, H., Czernilofsky, A. P., Purchio, A. F., Erikson, R. L., Bishop, J. M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6013-6017.
27. Honegger, A., Dull, T. J., Bellot, F., Van Obberghen, E., Szapary, D., Schmidt, A., Ullrich, A., and Schlessinger, J. (1988) *EMBO J.* **7**, 3045-3052.
28. Schuh, S. M., and Brugge, J. S. (1988) *Mol. Cell. Biol.* **8**, 2465-2471.
29. Courtneidge, S., and Smith, A. E. (1984) *EMBO J.* **3**, 585-591.
30. Bolen, J. B., Thiele, C. J., Israel, M. A., Yonemoto, W., Lipsich, L. A., and Brugge, J. S. (1984) *Cell* **38**,767-777.
31. Schaffhausen, B., and Benjamin, T. (1981) *J. Virol.* **40**, 184-196.
32. Luttrel, D. K., Luttrel, L. M., and Parsons, S. J. (1988) *Mol. Cell. Biol.* **8**, 497-501.