58. M. L. Alexander, M. A. Johnson, W. C. Lineberger, *J. Chem. Phys.* **82**, 5288 (1985).
59. D. E. Hunton, M. Hofmann, T. G. Lindeman, C. R. Albertoni, A. W. Castleman, Jr., *ibid.*, p. 2884.
60. W. D. Reents, Jr., A. M. Mujsce, V. E. Bondybey, M. L. Mandich, *ibid.* **86**, 5568 (1987).
61. M. D. Morse, M. E. Geusic, J. R. Heath, R. E. Smalley, *ibid.* **83**, 2293 (1985).
62. E. K. Parks, K. Liu, S. C. Richtsmeier, L. G. Pobo, S. J. Riley, *ibid.* **82**, 5470 (1985).
63. S. A. Ruatta, L. Hanley, S. L. Anderson, *Chem. Phys. Lett.* **137**, 5 (1987).
64. R. D. Levine, R. B. Bernstein, Eds., *Molecular Reaction Dynamics and Chemical Reactivity* (Oxford Univ. Press, New York, 1987).
65. B. Koplitz *et al.*, *Faraday Discuss. Chem. Soc.* **82**, 125 (1986).
66. N. F. Scherer, L. R. Khundkar, R. B. Bernstein, A. H. Zewail, *J. Chem. Phys.* **87**, 1451 (1987).
67. S. Morgan and A. W. Castleman, Jr., *J. Am. Chem. Soc.* **109**, 2867 (1987).
68. D. K. Bohme, *NATO ASI Ser. C* **118**, 111 (1984).
69. Financial support by the Army Research Office, grant DAAG29-85-K-0215, the Department of Energy, grant DE-AC02-82ER60055, and the National Science Foundation, grant ATM-82-04010, is gratefully acknowledged.

The Protein Kinase Family: Conserved Features and Deduced Phylogeny of the Catalytic Domains

STEVEN K. HANKS, ANNE MARIE QUINN, TONY HUNTER

In recent years, members of the protein kinase family have been discovered at an accelerated pace. Most were first described, not through the traditional biochemical approach of protein purification and enzyme assay, but as putative protein kinase amino acid sequences deduced from the nucleotide sequences of molecularly cloned genes or complementary DNAs. Phylogenetic mapping of the conserved protein kinase catalytic domains can serve as a useful first step in the functional characterization of these newly identified family members.

THE PROTEIN KINASES ARE A LARGE FAMILY OF ENZYMES, many of which mediate the response of eukaryotic cells to external stimuli (*1, 2*). The number of unique members of the protein kinase family that have been described has recently risen exponentially (*3*) and now approaches 100. The surge in the number of known protein kinases has been largely due to the advent of gene cloning and sequencing techniques. Amino acid sequences deduced from nucleotide sequences are considered to represent protein kinases if they include certain key residues that are highly conserved in the protein kinase "catalytic domain."

Two different molecular approaches have been most instrumental in the isolation of novel protein kinase–encoding genes or cDNAs: (i) complementation or suppression of genetic defects in invertebrate regulatory mutants, and (ii) screening DNA libraries by using protein kinase genes as hybridization probes under low stringency conditions. Recently, an approach that uses degenerate oligonucleotides as probes has led to the identification of several novel putative protein kinase genes and cDNAs (*4, 5*). The oligonucleotide probes are designed to recognize target sequences that encode short amino acid stretches highly conserved in protein kinase catalytic domains.

In this article, we present an alignment of catalytic domain amino acid sequences from 65 different members of the protein kinase family, including many putative protein kinase sequences recently deduced from nucleotide sequence data. Based on this alignment, we first identify and discuss conserved features of the catalytic domains and then provide a visual display of the various intersequence relations through construction of a catalytic domain phylogenetic tree. Catalytic domains from protein kinases having similar modes of regulation or substrate specificities are found to cluster together within the tree. This clustering would appear to be of predictive value in the determination of the properties and function of novel protein kinases.

Catalytic Domain Amino Acid Sequences

Protein kinase catalytic domains range from 250 to 300 amino acid residues, corresponding to about 30 kD. Fairly precise boundaries for the catalytic domains have been defined through an analysis of conserved sequences (*1, 6*, see below) as well as by assay of truncated enzymes (*7, 8*). The location of the catalytic domain within the protein is not fixed but, in most single subunit enzymes it lies near the carboxyl terminus, the amino terminus being devoted to a regulatory role. In protein kinases having a multiple subunit structure, subunit polypeptides consisting almost entirely of catalytic domain are common. All protein kinases thus far characterized with regard to substrate specificity fall within one of two broad classes, serine/threonine–specific and tyrosine-specific. Although both classes of protein kinase have very similar catalytic domain primary structures, certain short amino acid stretches appear to characterize each class (*4*), and these regions can be used to predict whether a putative protein kinase will phosphorylate tyrosine or serine/threonine.

Members of the protein–serine/threonine kinase and protein-

S. K. Hanks is a senior research associate at the Molecular Biology Laboratory, Salk Institute for Biological Studies, Post Office Box 85800, San Diego, CA 92138. A. M. Quinn is a scientific applications programmer at the Biocomputing Center, Salk Institute for Biological Studies, Post Office Box 85800, San Diego, CA 92138. T. Hunter is a professor at the Molecular Biology and Virology Laboratory, Salk Institute for Biological Studies, Post Office Box 85800, San Diego, CA 92138.

tyrosine kinase families with reported catalytic domain amino acid sequences are listed in Tables 1 and 2, respectively. They are classified within the tables according to similarities in primary structure, based on deduced catalytic domain phylogeny. Included in the tables are all confirmed and putative protein kinases for which the catalytic domain sequence was available as of November 1987 (*9*). Presumed functional homologs from different vertebrate species are listed together. Presumed invertebrate functional homologs of protein kinases also found in vertebrates, however, are given separate listings as a reflection of greater evolutionary distance and the possibility of functional divergence. The asterisks indicate protein kinases that have catalytic domains that are included in the amino acid sequence alignment. We will use the abbreviated names from the tables to refer to individual protein kinases.

Of the 45 unique vertebrate protein kinase family members included in Tables 1 and 2, 22 are serine/threonine–specific and 23 are tyrosine-specific. Fourteen of the vertebrate protein–serine/threonine kinases fall within one of the three subgroups that can be

Table 1. Protein–serine/threonine kinase family members.

A. Cyclic nucleotide–dependent subfamily
- cAPK-α: cAMP-dependent protein kinase catalytic subunit, α form
  - *-bovine cardiac muscle protein (*26*)
  - -mouse S49 lymphoma cell cDNA (*35*)
- cAPK-β: cAMP-dependent protein kinase catalytic subunit, β form
  - *-bovine pituitary cDNA (*36*)
  - -mouse S49 lymphoma cell cDNA (*37*)
- SRA3: cAMP-dependent protein kinase from yeast, RAS suppressor
  - *-*Saccharomyces cerevisiae* genomic DNA (*38*)
- TPK1(PK25): cAMP-dependent protein kinase from yeast, type 1
  - *-*S. cerevisiae* genomic DNA (*39, 40*)
- TPK2: cAMP-dependent protein kinase from yeast, type 2
  - *-*S. cerevisiae* genomic DNA (*39*)
- TPK3: cAMP-dependent protein kinase from yeast, type 3
  - *-*S. cerevisiae* genomic DNA (*39*)
- cGPK: guanosine 3′,5′-monophosphate (cGMP)–dependent protein kinase
  - *-bovine lung protein (*41*)

B. Calcium-phospholipid–dependent subfamily
- PKC-α: protein kinase C, α form
  - *-bovine brain cDNA (*42*)
  - -rabbit brain cDNA (*43*)
  - -human brain cDNA (partial) (*44*)
- PKC-β: protein kinase C, β form
  - *-bovine brain cDNA (*44*)
  - -rat brain cDNA (two splice forms) (*45, 46*)
  - -rabbit brain cDNA (two splice forms) (*43*)
  - -human brain cDNA (*44*)
- PKC-γ: protein kinase C, γ form
  - *-bovine brain cDNA (*44*)
  - -rat brain cDNA (*45*)
  - -human brain cDNA (*44*)
- PKC-ε: protein kinase C, ε form
  - -rat brain cDNA (RP16 clone) (partial) (*46*)
- DPKC: *Drosophila* gene product related to protein kinase C
  - *-*D. melanogaster* cDNA (*47*)

C. Calcium-calmodulin–dependent subfamily
- CaMII-α: calcium-calmodulin–dependent protein kinase type II, α subunit
  - *-rat brain cDNA (*48*)
- CaMII-β: calcium-calmodulin–dependent protein kinase type II, β subunit
  - *-rat brain cDNA (*49*)
- PhK-γ: phosphorylase kinase, γ subunit
  - *-rabbit skeletal muscle protein and cDNA (*50*)
  - -mouse muscle cDNA (*51*)
- MLCK-K: myosin light chain kinase, skeletal muscle
  - *-rabbit skeletal muscle protein (*52*)
- MLCK-M: myosin light chain kinase, smooth muscle
  - *-chicken gizzard cDNA (*53*)
- PSK-H1: putative protein-serine kinase
  - *-human HeLa cell cDNA (*4, 54*)
- PSK-C3: putative protein-serine kinase
  - -human HeLa cell cDNA (partial) (*4*)

D. SNF1 subfamily
- SNF1: "sucrose nonfermenting" mutant wild-type gene product
  - *-*S. cerevisiae* genomic DNA (*55*)
- nim1$^+$: "new inducer of mitosis"; suppressor of cdc25 mutants
  - *-*Schizosaccharomyces pombe* genomic DNA (*56*)
- KIN1: putative yeast protein kinase
  - *-*Saccharomyces cerevisiae* genomic DNA (*5*)
- KIN2: putative yeast protein kinase related to KIN1
  - *-*S. cerevisiae* genomic DNA (*5*)

E. CDC28-cdc2$^+$ subfamily
- CDC28: "cell-division-cycle" gene product in yeast
  - *-*S. cerevisiae* genomic DNA (*57*)
- cdc2$^+$: "cell-division-cycle" gene product in yeast
  - *-*Schizosaccharomyces pombe* genomic DNA (*58*)
- CDC2Hs: human functional homolog of cdc2$^+$
  - *-human transformed cell line cDNA (*33*)
- PSK-J3: putative protein kinase related to CDC28-cdc2$^+$
  - *-human HeLa cell cDNA (*4, 59*)
- KIN28: putative protein kinase related to CDC28-cdc2$^+$
  - *-*Saccharomyces cerevisiae* genomic DNA (*60*)

F. Casein kinase subfamily
- CKIIα: casein kinase II, α subunit
  - -bovine lung protein (partial) (*61*)
- DCKII: *Drosophila* casein kinase II, α subunit
  - *-*D. melanogaster* cDNA (*62*)

G. Raf-Mos proto-oncogene subfamily
- Raf: cellular homolog of oncogene products from 3611 murine sarcoma virus and Mill Hill 2 avian acute leukemia virus
  - *-human fetal liver cDNA (*63*)
- A-Raf: cellular oncogene product closely related to Raf
  - *-human T cell cDNA (*64*)
  - -mouse spleen cDNA (*65*)
- PKS: cellular gene product closely related to Raf
  - *-human fetal liver cDNA (*66*)
- Mos: cellular homolog of oncogene product from Moloney murine sarcoma virus
  - *-human placenta genomic DNA (*67*)
  - -mouse NIH 3T3 cell genomic DNA (*68*)
  - -rat 3Y1 cell genomic DNA (*69*)

H. STE7 subfamily
- STE7: "sterile" mutant wild-type allele gene product
  - *-*S. cerevisiae* genomic DNA (*70*)
- PBS2: polymixin B antibiotic resistance gene product
  - *-*S. cerevisiae* genomic DNA (*71*)

I. Family members with no close relatives
- CDC7: "cell-division-cycle" gene product
  - *-*S. cerevisiae* genomic DNA (*72*)
- weel$^+$: "reduced size at division" mutant wild-type gene product
  - *-*Schizosaccharomyces pombe* genomic DNA (*73*)
- ran1$^+$: "meiotic bypass" mutant wild-type allele gene product
  - *-*S. pombe* genomic DNA (*74*)
- PIM-1: putative transforming protein induced by murine leukemia virus integration
  - *-mouse BALB/c cell genomic DNA (*75*)
- HSVK: herpes simplex virus-US3 gene product
  - *-herpes simplex virus genomic DNA (*76*)

*Protein kinases that have catalytic domains included in the amino acid sequence alignment.

classified according to their mode of regulation: cyclic nucleotide–dependent, calcium-phospholipid–dependent, and calcium-calmodulin–dependent. Two of the serine/threonine kinases, Mos and Raf (products of the c-*mos* and c-*raf* genes, respectively), are cellular homologs of transforming proteins encoded by the retroviral oncogenes. Other members of the serine/threonine group with demonstrated oncogenic potential are A-Raf (a distinct Raf-related member), and PIM-1 (a putative transforming protein activated by viral integration). Three vertebrate serine/threonine kinases (CDC2Hs, PSK-J3, and CKIIα) are closely related, by various degrees, to the yeast cell cycle control protein kinases CDC28 and $cdc2^+$. A protein–serine/threonine kinase has been described in herpes simplex virus (HSVK) and, like the retroviral oncogenes, probably originated as a eukaryotic cellular sequence. The protein-tyrosine kinases can be further grouped as members of either the Src subfamily or one of three different growth factor receptor subfamilies. The protein-tyrosine kinases encoded by the c-*abl* and c-*fes/fps* genes may be considered distant members of the Src subfamily. At least nine of the protein-tyrosine kinase genes have been transduced by retroviruses where they encode transforming proteins.

Twenty-five additional sequences listed in Tables 1 and 2 derive from invertebrate species. Eight are from *Drosophila*, one from nematode, and the other 16 are from the budding or fission yeasts. Many of the *Drosophila* protein kinases, as well as the nematode protein kinase, were identified by screening DNA libraries with probes from a vertebrate protein kinase gene or cDNA and thus are likely to represent functional homologs of the vertebrate enzymes. The *Drosophila* "sevenless" (7less) protein kinase and most of the yeast protein kinases were identified through molecular genetics. All of the yeast protein kinases identified to date fall within the serine/threonine–specific class, despite directed attempts to identify protein-tyrosine kinases in yeast (*5*). This observation, together with the fact that many of the protein-tyrosine kinase catalytic domains are components of growth factor receptor molecules, suggests that tyrosine specificity may have been a recent development in catalytic domain evolution, arising in conjunction with the acquisition of multicellularity and serving a role in cell-cell communication.

Table 2. Protein-tyrosine kinase family members.

A. Src subfamily
- Src: cellular homolog of oncogene product from Rous avian sarcoma virus
  - *-human fetal liver genomic DNA (*77*)
  - -mouse brain cDNA; neuronal alternate splice form (*78*)
  - -chicken genomic DNA (*79*)
  - -*Xenopus laevis* ovary cDNA (partial) (*80*)
- Yes: cellular homolog of oncogene product from Yamaguchi 73 avian sarcoma virus
  - *-human embryo fibroblast cDNA (*81*)
- Fgr: cellular homolog of oncogene product from Gardner-Rasheed feline sarcoma virus
  - *-human genomic DNA (*82*)
  - -human B lymphocyte cell line cDNA (amino terminus) (*83*)
- FYN: putative protein-tyrosine kinase related to Fgr and Yes
  - *-human fibroblast cDNA (*84*)
- LYN: putative protein-tyrosine kinase related to LCK and Yes
  - *-human placenta cDNA (*85*)
- LCK: lymphoid cell protein-tyrosine kinase
  - *-human (JURKAT) T cell leukemia line cDNA (*86*)
  - -mouse (LSTRA) T cell lymphoma line cDNA (*87*)
- HCK: hematopoietic cell putative protein-tyrosine kinase
  - *-human placenta and peripheral leukocyte cDNAs (*88*)
- Dsrc64: *Drosophila* gene product related to Src; polytene locus 64B
  - *-*D. melanogaster* genomic DNA (*89, 90*)
- Dsrc28: *Drosophila* gene product related to Src; polytene locus 28C
  - *-*D. melanogaster* adult female cDNA (*91*)

B. Abl subfamily
- Abl: cellular homolog of oncogene product from Abelson murine leukemia virus
  - *-human fetal liver cDNA (*92*)
- ARG: putative protein-tyrosine kinase related to Abl
  - -human genomic DNA (partial) (*93*)
- Dash: *Drosophila* gene product related to Abl
  - *-*D. melanogaster* genomic DNA (*90*)
- Nabl: nematode gene product related to Abl
  - *-*Caenorhabditis elegans* genomic DNA (*94*)
- Fes/Fps: cellular homolog of oncogene products from Gardner-Arnstein and Snyder-Theilen feline sarcoma viruses and Fujinami and PRCII avian sarcoma viruses
  - *-human genomic DNA (*95*)
  - -feline genomic DNA (*96*)
  - -chicken genomic DNA (*97*)

C. Epidermal growth factor receptor subfamily
- EGFR: epidermal growth factor receptor; cellular homolog of oncogene product (v-Erb-B) from AEV-H avian erythroblastosis virus
  - *-human placenta and A431 cell line cDNAs (*98*)
- NEU: cellular oncogene product activated in induced rat neuroblastomas (also called ERB-B2 or HER2)
  - *-human placenta and gastric cancer cell line cDNAs (*99*)
  - -rat neuroblastoma cell line cDNA (*100*)
- DER: *Drosophila* gene product related to EGFR
  - *-*D. melanogaster* genomic DNA (*101*)

D. Insulin receptor subfamily
- INS.R: insulin receptor
  - *-human placenta cDNA (*102*)
- IGF1R: insulin-like growth factor 1 receptor
  - *-human placenta cDNA (*103*)
- DILR: *Drosophila* gene product related to INS.R
  - *-*D. melanogaster* embryo cDNA (*104*)
- Ros: cellular homolog of oncogene product from UR2 avian sarcoma virus
  - *-human placenta genomic DNA (*105*)
  - -chicken genomic DNA (*106*), chicken kidney cDNA (*107*)
- 7less: *Drosophila sevenless* gene product essential for R7 photoreceptor cell development
  - *-*D. melanogaster* eye imaginal disc cDNA (*108*)
- TRK: colon carcinoma oncogene product activated by genetic recombination
  - *-human tumor cell cDNA (*109*)
- MET: *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine (MNNG)–induced oncogene product
  - *-human HOS cell line cDNA (*110*)

E. Platelet-derived growth factor receptor subfamily
- PDGFR: platelet-derived growth factor receptor
  - *-mouse NR6 fibroblast cell line cDNA (*111*)
- CSF1R: colony-stimulating factor–type 1 receptor; cellular homolog of oncogene product (v-Fms) from McDonough feline sarcoma virus
  - *-human placenta cDNA (*112*)
- Kit: cellular homolog of oncogene product from Hardy-Zuckerman 4 feline sarcoma virus
  - *-human placenta cDNA (*113*)
- RET: cellular oncogene product activated by recombination
  - *-human T cell lymphoma cDNA (*114*)

F. Other receptor-like protein-tyrosine kinases
- TKR11: putative protein-tyrosine kinase
  - -chicken genomic DNA (partial) (*115*)
- TKR16: putative protein-tyrosine kinase
  - -chicken genomic DNA (partial) (*115*)

*Protein kinases that have catalytic domains included in the amino acid sequence alignment.

Conserved Features of the Catalytic Domains

To compare primary structures, we have aligned catalytic domains from the 65 protein kinases marked by an asterisk in Tables 1 and 2 (Fig. 1). The 65 sequences represent each of the separate entries in the Tables except for six family members that are not included because their catalytic domain sequences have been only partially determined. The alignment was made by eye and is parsimonious in nature; the amount of gapping introduced into the sequences in order to optimize positional similarities was kept to a minimum. The alignment clearly demonstrates the overall similarity among the catalytic domains. The catalytic domains are not conserved uniformly but, rather, consist of alternating regions of high and low conservation. Eleven major conserved subdomains are evident (Fig. 1, I to XI), separated by regions of lower conservation wherein fall the larger gaps or inserts. Very large inserts (in excess of 60 residues) occur in CDC7 between subdomains VII and VIII and between subdomains X and XI, and in PDGFR, CSF1R, and Kit between subdomains V and VI. A similarity profile of the aligned catalytic domains provides a ready visualization of the subdomain structure (Fig. 2). Such an arrangement of alternating regions of high and low conservation is a common feature of homologous globular proteins (*10*) and gives some clues to higher order structure. The conserved subdomains must be important for catalytic function, either directly as components of the active site or indirectly by contributing to the formation of the active site through constraints imposed on secondary structure. The nonconserved regions, on the other hand, are likely to occur in loop structures, where folding allows the essential conserved regions to come together.

Highly conserved individual amino acids within the catalytic domains are expected to play important roles in catalysis. We will refer to amino acid positions using the residue numbering for bovine adenosine 3′,5′-monophosphate (cAMP)–dependent protein kinase catalytic subunit, α form (cAPK-α, Fig. 1). Nine positions in the alignment contain the identical amino acid residue in each of the 65 sequences. These invariant residues correspond to cAPK-α: Gly^{52}, Lys^{72}, Glu^{91}, Asp^{166}, Asn^{171}, Asp^{184}, Gly^{186}, Glu^{208}, and Arg^{280}. An additional five positions contain the identical amino acid in all but one of the sequences: Gly^{50}, Val^{157}, Phe^{185}, Asp^{220}, and Gly^{225}. Many of these most highly conserved residues directly participate in adenosine triphosphate (ATP) binding and phosphotransfer.

The consensus Gly-X-Gly-X-X-Gly, found in many nucleotide binding proteins in addition to the protein kinases (*11*), is found in subdomain I, very near the catalytic domain amino terminus. The invariant or nearly invariant residues corresponding to cAPK-α Gly^{50} and Gly^{52} fall within this consensus. Only two positions on the amino-terminal side of this consensus show conservation throughout the protein kinase family; hydrophobic residues occupy positions one and seven upstream from the first glycine in the consensus. The amino terminus of some catalytic domain polypeptides lies as close as ten residues from the first conserved glycine. A model for the ATP-binding site of v-Src (*12*), based on the three-dimensional structures from other nucleotide binding proteins, shows the Gly-X-Gly-X-X-Gly residues forming an elbow around the nucleotide, with the first glycine in contact with the ribose moiety and the second glycine lying near the terminal pyrophosphate. A nearly invariant valine residue lies within subdomain I, located just two positions on the carboxyl-terminal side of the Gly-X-Gly-X-X-Gly consensus (Val^{57} for cAPK-α) and may contribute to the positioning of the conserved glycines.

In subdomain II lies an invariant lysine, corresponding to cAPK-α Lys^{72}, that is certainly the best characterized catalytic domain residue. This lysine appears to be directly involved in the phosphotransfer reaction, possibly mediating proton transfer (*13*). In cAPK-α (*14*), v-Src (*15*), and EGFR (*16*), Lys^{72} or its equivalent reacts with the ATP analog *p*-fluorosulfonyl 5′-benzoyl adenosine, thereby inhibiting enzyme activity. Site-directed mutagenesis techniques have been used to substitute alternate amino acids at this position in v-Src (*13, 17*), v-Mos (*18*), v-Fps (*19*), EGFR (*20*), INS.R (*21*), and PDGFR (*22*). All substitutions, including arginine, result in loss of protein kinase activity. In all but three of the aligned sequences, an alanine is present two positions on the amino-terminal side of the invariant lysine in subdomain II. The invariant lysine lies 14 to 23 residues downstream of the last conserved glycine in subdomain I, but no mutations have been made to test whether this spacing is critical.

The central core of the catalytic domain, the region with greatest frequency of highly conserved residues, consists of subdomains VI through IX. The invariant or nearly invariant residues in subdomain VI (corresponding to Asp^{166} and Asn^{171}) and subdomain VII (corresponding to Asp^{184}, Phe^{185}, and Gly^{186}) also have been implicated in ATP binding. These residues are part of a feature found in a number of bacterial phosphotransferases that use ATP as phosphate donor (*23*). The aspartic acid residues corresponding to cAPK-α Asp^{166} and Asp^{184} may interact with the phosphate groups of ATP through Mg^{2+} salt bridges (*23*). The triplet corresponding to Asp^{184}-Phe^{185}-Gly^{186} in subdomain VII is of further interest in that it represents the most highly conserved short stretch in the catalytic domains. It is flanked for two positions on either side by hydrophobic or near-neutral residues.

Subdomain VIII contains the consensus triplet Ala-Pro-Glu, a conserved feature often mentioned as a key protein kinase catalytic domain indicator (*1*). The invariant residue corresponding to cAPK-α Glu^{208} contributes to the Ala-Pro-Glu consensus. In addition to the conservation of these residues, several other lines of evidence implicate this region as important in catalysis. Mutagenesis studies have shown that each residue in the Ala-Pro-Glu consensus is required for activity of v-Src (*24*). Other studies have provided evidence that this consensus lies very near the catalytic site. An affinity peptide substrate analog reacts with cAPK-α Cys^{199}, thereby inhibiting enzyme activity (*25*). Also, sites of autophosphorylation found in many protein-tyrosine kinases (*1*) as well as cAMP-dependent protein kinase [Thr^{197} (*26*)] lie within 20 residues upstream of the Ala-Pro-Glu consensus. The role of this autophosphorylation site is not entirely settled, but for several protein-tyrosine kinases there is evidence that phosphorylation of this site leads to increased catalytic activity (*27*). Autophosphorylation may result in a conformational change that allows better access of exogenous substrates to the active site.

Subdomains VI and VIII are of additional interest in that they contain residues that are specifically conserved in either the protein–serine/threonine or the protein-tyrosine kinases and, as such, may play a role in recognition of the correct hydroxyamino acid. The most striking indicator of amino acid specificity is found in subdomain VI, lying between the invariant residues corresponding to cAPK-α Asp^{166} and Asn^{171}; two of the residues implicated in ATP binding. The consensus Asp-Leu-Lys-Pro-Glu-Asn in this region is a strong indicator of serine/threonine specificity, whereas the protein-tyrosine kinase consensus is either Asp-Leu-Arg-Ala-Ala-Asn (for the vertebrate members of the Src subfamily) or Asp-Leu-Ala-Ala-Arg-Asn (for all others). Another such region is found in subdomain VIII and lies immediately on the amino-terminal side of the Ala-Pro-Glu consensus. This region is highly conserved among the protein-tyrosine kinases with a more limited conservation among the protein–serine/threonine kinases. The protein-tyrosine kinase consensus through this region is Pro-Ile/Val-Lys/Arg-Trp-Thr/Met-Ala-Pro-Glu while the protein–serine/threonine kinase

consensus is Gly-Thr/Ser-X-X-Tyr/Phe-X-Ala-Pro-Glu. These regions in subdomains VI and VIII that indicate substrate specificity have been targeted for the design of degenerate oligonucleotide probes for use in screening cDNA libraries to identify novel members of both the protein–serine/threonine (*4*) and protein-tyrosine (*28*) kinase families.

To date, no evidence has been reported concerning the possible functions of residues in conserved subdomains III, IV, V, IX, X, and XI. Subdomain IX contains a very well conserved short stretch that includes the nearly invariant residues corresponding to Asp^{220} and Gly^{225}. Subdomains III and XI each contain an invariant residue, corresponding to Glu^{91} and Arg^{280}. The latter or its equivalent must lie very near the catalytic domain carboxyl terminus. Arginine residues occupying this position reside just 16 residues upstream from both the CDC28 and HSVK polypeptide carboxyl termini, and just 19 residues upstream from both the Mos and Fes carboxyl termini. Deletion analysis of v-Src places the carboxyl terminus of the catalytic domain of the protein-tyrosine kinases at a conserved hydrophobic residue ten residues downstream of this arginine (*8*). The point mutation conferring temperature sensitivity in some *cdc28* mutants replaces this conserved arginine with glutamine (*29*).

A leap in our understanding of the functional roles of the conserved catalytic domain residues will come with the solution of a crystal structure for one of the protein kinase catalytic domains. The similarities in primary strucure should carry over to the higher order structure and catalytic mechanism as well. Other investigators have been making progress toward the solution of the three-dimensional structure of cAPK-α (*30*).

Catalytic Domain Phylogeny

Amino acid sequence alignments can be used to deduce phylogenetic relationships (*31*). We have used the alignment data from Fig. 1 to construct a phylogenetic tree of the protein kinase catalytic domains (Fig. 3). All 65 of the sequences in the alignment are included in the tree. They derive from both vertebrate and invertebrate sources and, in some cases, presumed functional homologs from both vertebrate and invertebrate sources are represented. The tree, therefore, reflects catalytic domain evolution stemming from gene duplication events (for example, when the vertebrate, mostly human, sequences are compared), speciation events (when vertebrate and invertebrate functional homologs are compared), or both.

The tree reveals a relation between catalytic domain sequence and certain biochemical properties; catalytic domains from protein kinases having similar modes of regulation or substrate specificities tend also to have similar primary structures and cluster together within the tree. Five major branch clusters are present in the tree: (i) protein-tyrosine kinases, (ii) cyclic nucleotide– and calcium-phospholipid–dependent protein kinases, (iii) calcium-calmodulin–dependent protein kinases, (iv) protein kinases closely related to SNF1, and (v) protein kinases closely related to CDC28. These major clusters account for all but 12 of the 65 sequences included in the tree. Generally, a sequence found within one of these clusters shares in excess of 35% identical amino acids with each of the other sequences in the cluster, whereas the catalytic domain sequences that do not map within the same cluster have identities in the range of 20 to 25%.

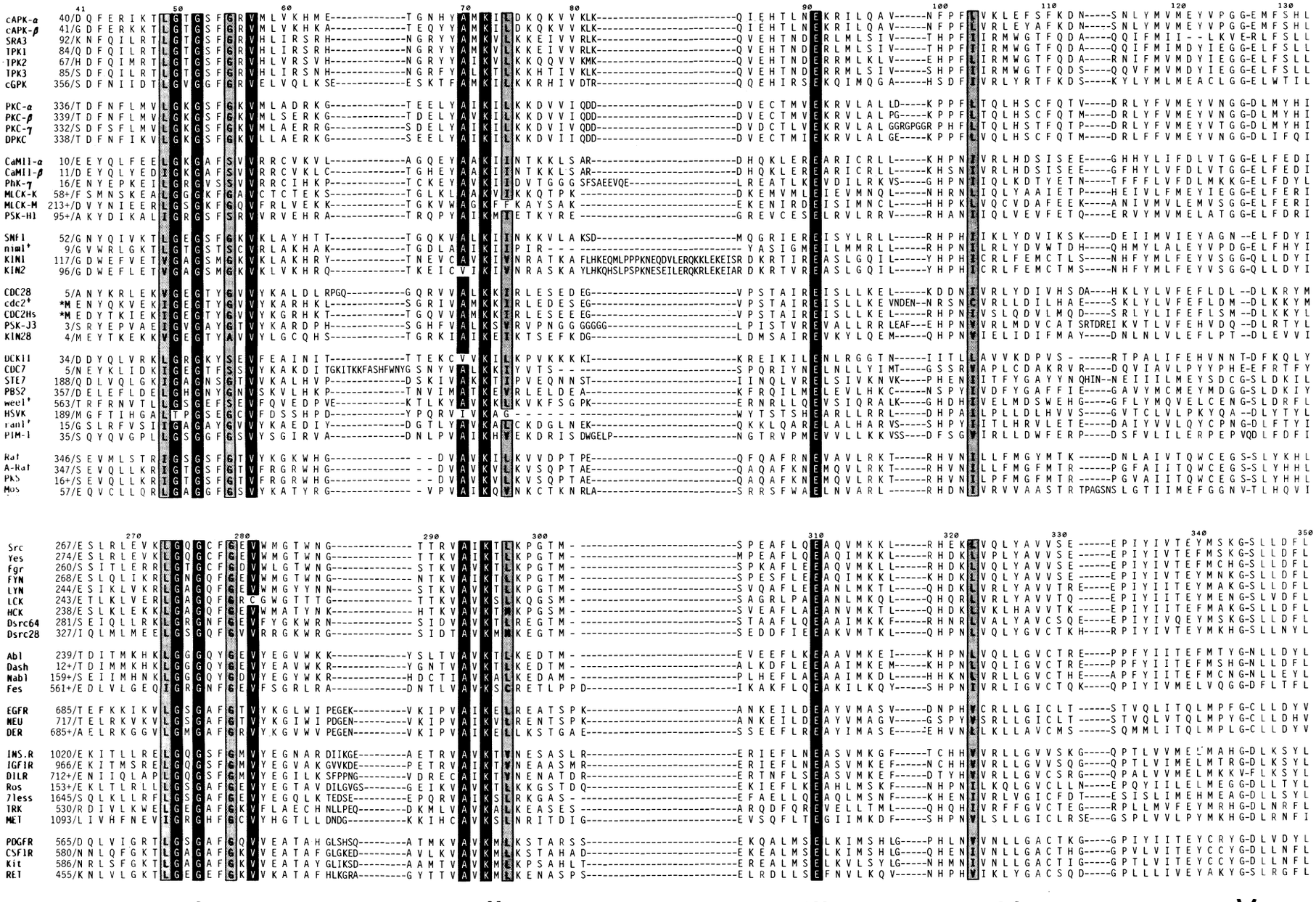
The most highly populated cluster contains all 27 confirmed or putative protein-tyrosine kinases. The large number of protein-tyrosine kinases probably reflects the intense research effort devoted to this group, rather than a true indication of their abundance relative to the protein–serine/threonine kinases. Branches leading to the Src subfamily and to each of the three receptor subfamilies diverge from the main line at about the same point. In light of the oncogenic potential of many of the protein-tyrosine kinases, it is of interest that the protein–serine/threonine kinases having the least divergence from this group include Raf and Mos, cellular homologs of retroviral oncogene products. However, another potentially oncogenic protein–serine/threonine kinase, PIM-1, is not closely related to the protein-tyrosine kinases.

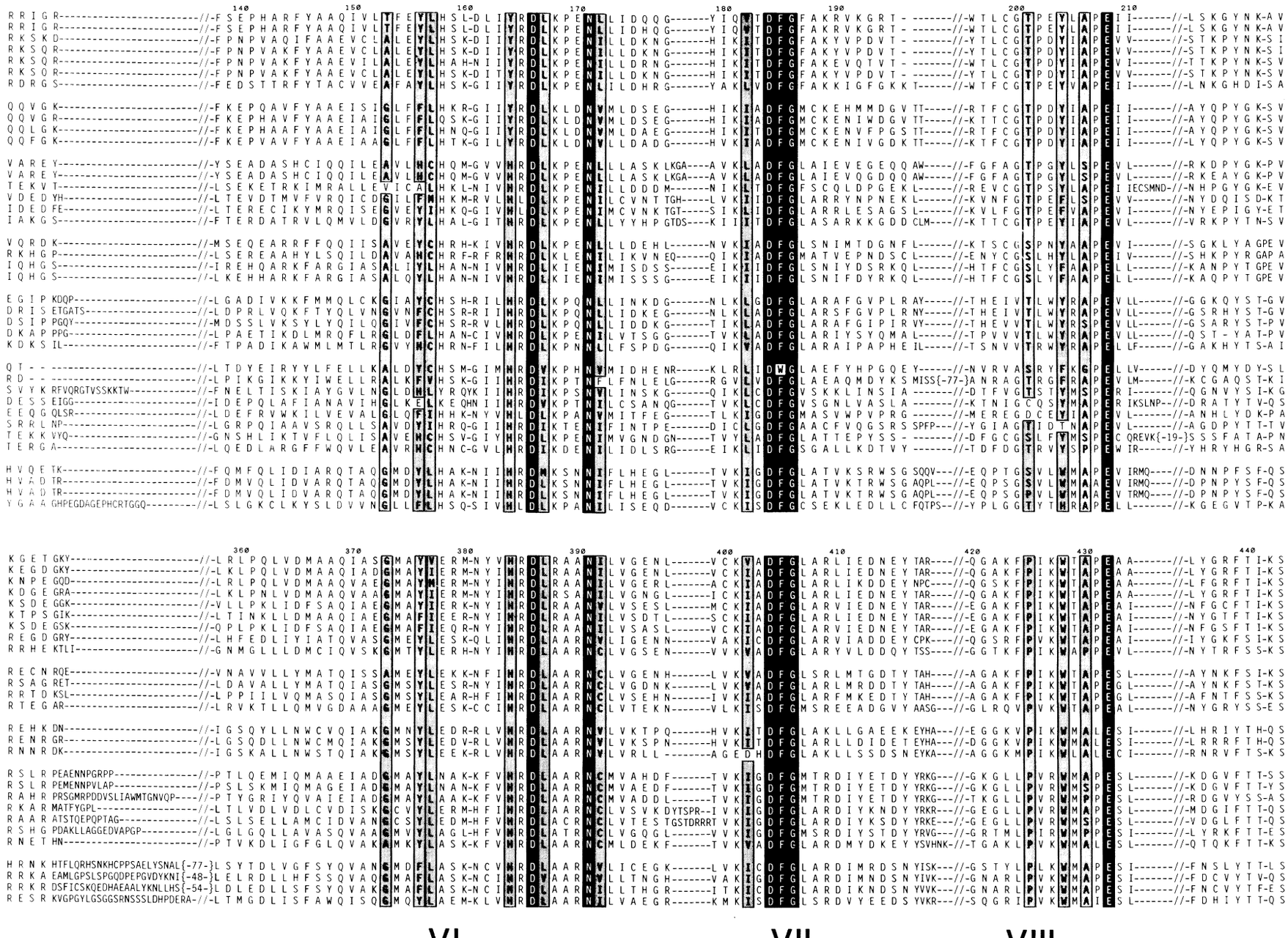
The next most populous cluster in the tree includes two separate subfamilies that can be classified according to their mode of regulation: the cyclic nucleotide–dependent protein kinases and the calcium-phospholipid–dependent protein kinases. The similarities in the mode of regulation of the members of these two subfamilies, namely, activation by "second messengers" released in response to ligand binding at the cell surface, may be a reflection of their recent evolutionary divergence.

The third major catalytic domain cluster contains the subfamily of protein kinases that have activities regulated by calmodulin. The calmodulin-dependent cluster falls near the cyclic nucleotide– and calcium-phospholipid–dependent cluster. All members of the calmodulin-dependent subfamily have a calmodulin binding domain, characterized by a high proportion of basic amino acid residues and having a propensity for formation of an amphiphilic α helix, residing outside the catalytic domain. (Note that the calmodulin binding domain sequences were not included in the phylogenetic analysis.) The different protein kinases thus far described as being regulated by calmodulin, therefore, appear to have diverged from a common ancestor after acquisition of the calmodulin binding domain. The mapping of the putative protein kinase PSK-H1 within this cluster predicts that this enzyme will also prove to be regulated by calmodulin.

Also mapping near the cyclic nucleotide– and calcium-phospholipid–dependent protein kinases is a small cluster composed of four protein kinases recently identified in the budding or fission yeasts; SNF1, $nim1^+$, KIN1, and KIN2. Whether these protein kinases

Fig. 1. Multiple amino acid sequence alignment of 65 protein kinase catalytic domains. The first 38 sequences derive from protein–serine/threonine kinases (indicated by asterisks in Table 1) and the remaining 27 sequences in the alignment are from protein-tyrosine kinases (indicated by asterisks in Table 2). cAPK-α and Src have been chosen as prototype protein–serine/threonine and protein-tyrosine kinases, respectively; their catalytic domain sequences are numbered to indicate residue position from the polypeptide amino terminus. (Although the human Src sequence is shown, the numbering is actually taken from the chicken Src sequence to maintain established convention). The number of additional amino- and carboxyl-terminal flanking residues lying outside the catalytic domains are shown at the beginning and end, respectively, of each sequence. In several cases the sequences have not been determined through to the polypeptide amino or carboxyl termini; for these, the number of determined residues is given followed by a plus (+) sign. An asterisk (*) at the beginning or end of a sequence indicates that no additional flanking residues are contained in the polypeptide. Gaps, represented by dashes, were introduced into the sequences to optimize the alignment. In six cases, long insert segments have been excluded from the alignment to shorten the figure. The positions and lengths of the excluded inserts within the alignment are indicated by numbers within braces (for example, {-48-}); the excluded gap positions in the other sequences that correspond to these long inserts are shown as double slashes (//). Residues conserved in 62 or more of the 65 sequences are shown as white letters in black boxes. Positions where residues of similar structure are conserved in 63 or more sequences are shown in shaded boxes. Structurally similar groupings used for this purpose are nonpolar chain R groups (M, L, I, V, and C); aromatic or ring-containing R groups; (F, Y, W, and H); small R groups with near neutral polarity (A, G, S, T, and P); acidic and uncharged polar R groups (D, E, N, and Q); and basic polar R groups (K, R, and H). The single-letter amino acid code is used (A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; and Y, tyrosine). Roman numerals at bottom indicate conserved subdomains.


I
II
III
IV
V


VI
VII
VIII

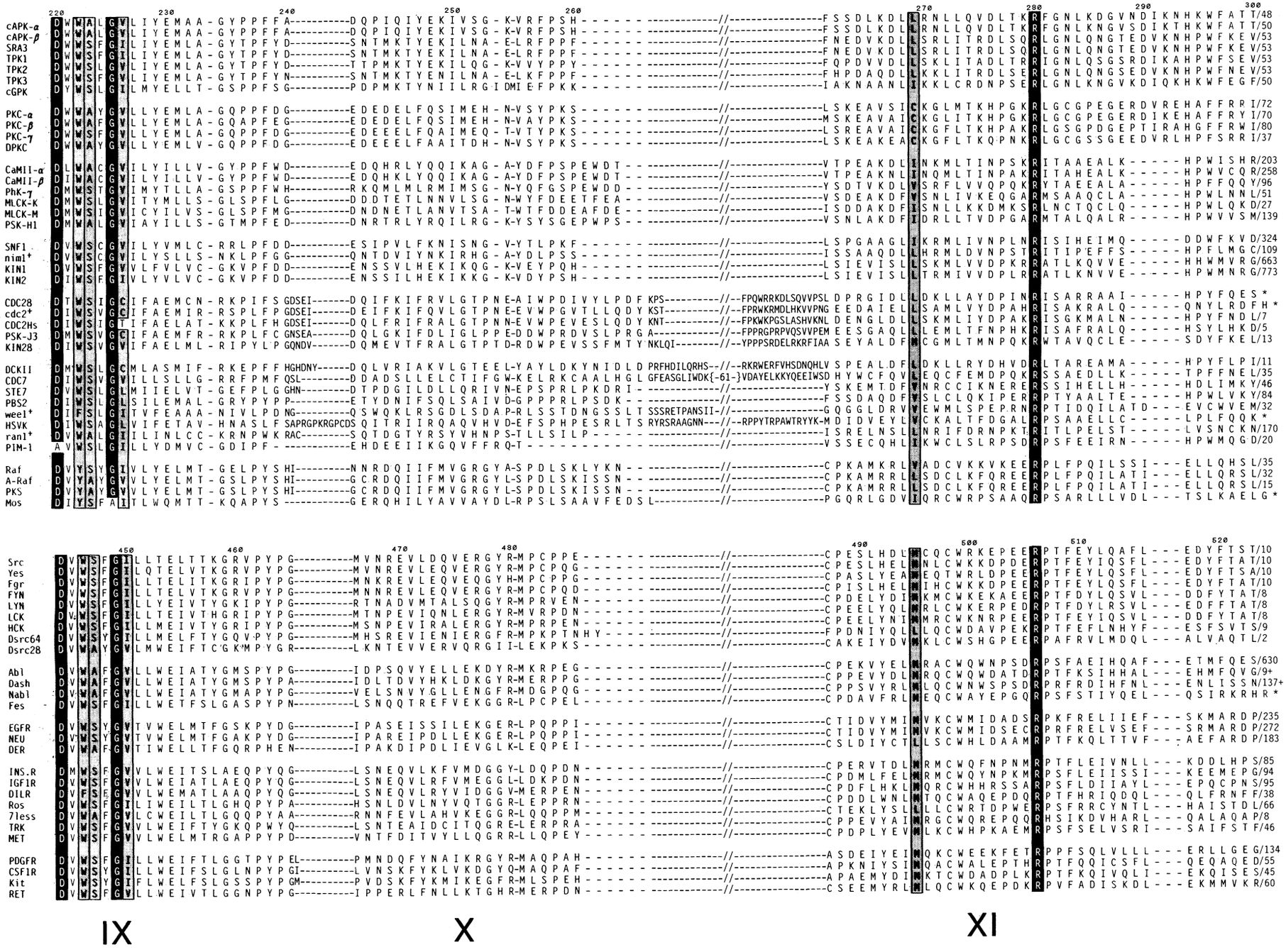
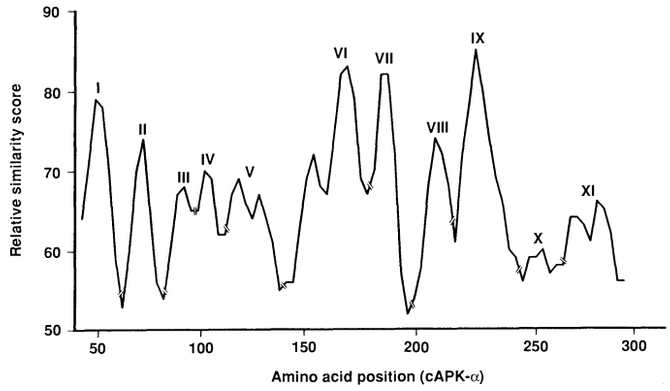
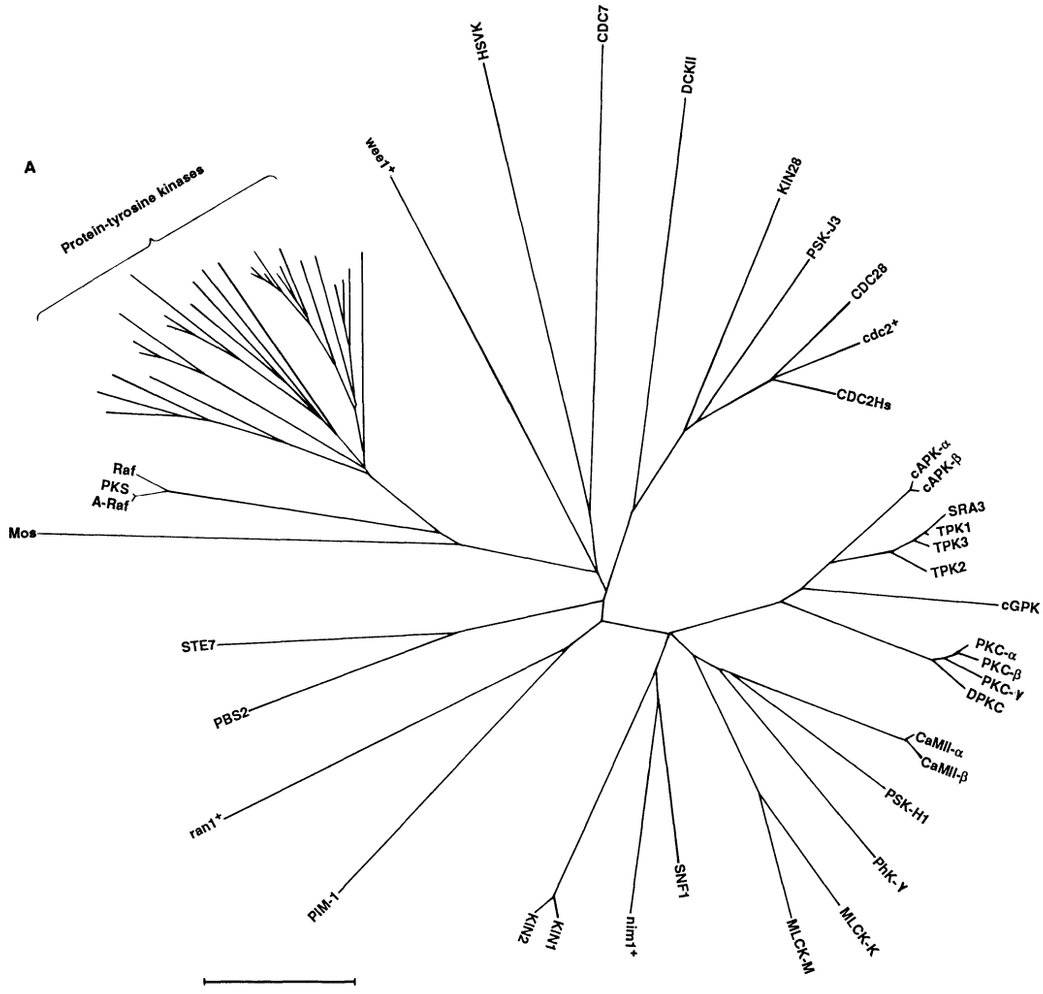

IX
X
XI

Fig. 2. Similarity profile of protein kinase catalytic domains. For each position in the alignment shown in Fig. 1, a relative similarity score was determined based on the "structure-genetic" scoring matrix (*116*) for amino acid similarities. Similarity scores were calculated as the sum of all possible pairwise comparisons between the individual amino acids at each position and expressed as the percentage of the highest possible score (that is, the score obtained when an identical residue occupies the position in all 65 aligned sequences). To smooth out the curve, a 9-position running average of the relative scores was determined, and every third position was plotted. Positions that contain gaps for ten or more of the sequences were not included in the profile; however, the locations of the major gap sites are indicated by breaks in the curve. The mean relative score for all the positions included in the profile is 66 with a standard deviation of 14.9. Relative similarity scores obtained when the catalytic domain sequences were randomly scrambled had a mean of 47 and standard deviation of 1.85. Roman numerals indicate conserved subdomains.





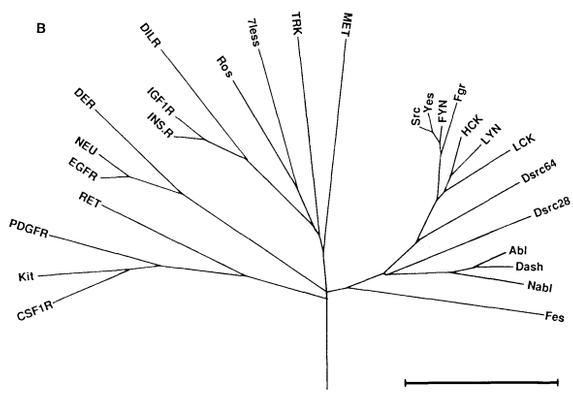


Fig. 3. Deduced phylogeny of protein kinase catalytic domains. The phylogenetic tree was constructed from the multiple alignment shown in Fig. 1. The tree-building concept of Fitch and Margoliash (*117*) was used as implemented by Feng and Doolittle (*118*). Briefly, similarity scores were obtained for all possible pairwise comparisons and transformed into a difference matrix from which branch order and length were determined. Programs were run on a VAX-785 computer equipped with 40 megabytes physical memory under virtual memory operating system (VMS). Systems limitations required that the branch lengths for the protein–serine/threonine and protein-tyrosine kinases be calculated separately, and the tree shown is thus a composite of these two determinations. The position of the protein-tyrosine kinase cluster was determined by including two protein-tyrosine kinases (Src and EGFR) in the protein–serine/threonine kinase tree construction. The individual sequences are indicated by the abbreviated names in Tables 1 and 2. The protein-tyrosine kinases are not labeled in (**A**), but are shown in the cluster enlargement in (**B**). The tree is shown "unrooted" in (A) as the branches are all measured relative to one another with no outside reference point. The scale bars represent a branch length corresponding to a relative difference score of 25. The tree depicted is likely to underestimate distances between the least related members of the family, particularly since the alignment used in its construction is parsimonious.

have similar modes of regulation remains to be determined. KIN1 and KIN2 were identified through screening a *Saccharomyces cerevisiae* DNA library with probes designed to recognize sequences characteristic of protein-tyrosine kinases and, as such, have been suggested to represent "structural mosaics" with some features of catalytic domain structure more indicative of the protein-tyrosine kinases than the protein–serine/threonine kinases (*5*). The deduced phylogeny of KIN1 and KIN2, however, does not suggest a close evolutionary relationship with protein-tyrosine kinases. In fact, the probe target used to identify KIN1 and KIN2 encodes the stretch of amino acids corresponding to cAPK-α Asp^{220}-Gly^{225} in conserved subdomain IX, a region of high conservation in all of the catalytic domains regardless of substrate specificity.

The subfamily related to CDC28 includes functional homologs from three widely divergent species: CDC28 from the budding yeast *S. cerevisiae*, $cdc2^+$ from the fission yeast *Schizosaccharomyces pombe*, and human CDC2Hs. Functional homology was demonstrated by heterologous complementation of conditional mutants defective in cell cycle progression (*32, 33*). The other two sequences mapping within this cluster are putative protein kinases identified in *Saccharomyces cerevisiae* (KIN28) and human HeLa cells (PSK-J3). The members of this cluster are also distinguished by the small sizes of the catalytic domain-containing polypeptides, suggesting their activities may be regulated through association with other polypeptides in a holoenzyme complex. Indeed, support for this notion has been obtained for $cdc2^+$ (*34*).

Perspectives

The tremendous diversity of the protein kinase family is just now beginning to be appreciated. Most of the catalytic domain sequences referenced in Tables 1 and 2 were reported within the past 2 years. With continued characterizations of regulatory mutants in invertebrates, along with the recent development of new hybridization approaches for the identification of DNA clones that encode novel protein kinase catalytic domains, it is likely that the rate of discovery will continue to accelerate through the next several years. The difficult tasks will be to confirm protein kinase activities for the newly identified family members and to elucidate their functional roles. Clues to function may come through an analysis of catalytic domain primary structure and subsequent phylogenetic mapping. A catalytic domain that has only limited divergence from another, better characterized, member of the family can be expected to play a similar role in cellular physiology. Further clues are likely to come from an inspection of amino acid sequences lying outside the catalytic domain where residues involved in enzyme regulation may be found.

REFERENCES AND NOTES

1. T. Hunter and J. A. Cooper, in *The Enzymes*, P. D. Boyer and E. G. Krebs, Eds. (Academic Press, Orlando, FL, 1986), vol. 17, pp. 191–246.
2. A. M. Edelman, D. K. Blumenthal, E. G. Krebs, *Annu. Rev. Biochem.* **56**, 567 (1987).
3. T. Hunter, *Cell* **50**, 823 (1987).
4. S. K. Hanks, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 388 (1987).
5. D. E. Levin, C. I. Hammond, R. O. Ralston, J. M. Bishop, *ibid.*, p. 6035.
6. W. C. Barker and M. O. Dayhoff, *ibid.* **79**, 2836 (1982).
7. A. D. Levinson, S. A. Courtneidge, J. M. Bishop, *ibid.* **78**, 1624 (1981); J. S. Brugge and D. Darrow, *J. Biol. Chem.* **259**, 4550 (1984); J. Y. J. Wang and D. Baltimore, *ibid.* **260**, 64 (1985); R. J. Bold and D. J. Donoghue, *Mol. Cell. Biol.* **5**, 3131 (1985); I. Sadowski and T. Pawson, *Oncogene* **1**, 181 (1987).
8. V. W. Wilkerson, D. L. Bryant, J. T. Parsons, *J. Virol.* **55**, 314 (1985); P. Yaciuk and D. Shalloway, *Mol. Cell. Biol.* **6**, 2807 (1986).
9. We have attempted to include all catalytic domain sequence reports except those encoded by retroviruses. Reports of partial sequences are included only if a complete catalytic domain sequence is not available. We have chosen not to include retroviral oncogene product sequences since the catalytic domains from this group are effectively represented by their closely related cellular counterparts. A listing of references for retroviral protein kinase sequences can be found in *Molecular Biology of Tumor Viruses: RNA Tumor Viruses* [R. A. Weiss, N. Teich, H. Varmus, J. Coffin, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, ed. 2, 1985)]. Since the time these tables were compiled, complete sequences for six additional members of the protein kinase family have been published: (i) EPH, a novel receptor-like protein-tyrosine kinase [H. Hirai, Y. Maru, K. Hagiwara, J. Nishida, F. Takaku, *Science* **238**, 1717 (1987)]; (ii) TKL, a novel member of the Src subfamily [K. Strebhardt, J. I. Mullins, C. Bruck, H. Rübsamen-Waigmann, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8778 (1987)]; (iii) *ninaC* protein, a *Drosophila* gene product essential for normal photoreceptor cell function [C. Montell and G. M. Rubin, *Cell* **52**, 757 (1988)]; (iv) *nim*A, a cell cycle control protein kinase from *Aspergillus* [S. A. Osmani, R. T. Pu, N. R. Morris, *Cell* **53**, 237 (1988)]. Further, in *Drosophila*, a full-length sequence of cAMP- and a partial sequence of cGMP-dependent protein kinase catalytic domain have been reported [J. L. Foster, G. C. Higgins, F. R. Jackson, *J. Biol. Chem.* **263**, 1676 (1988)] as have catalytic domain sequences from a number of vertebrate protein kinases previously reported from other vertebrate species (not cited); (v) $byr1^+$, a suppressor of sporulation defects in *Schizosaccharomyces pombe* [S. A. Nadin-Davis and A. Nasim, *EMBO J.* **7**, 985 (1988)]; and (vi) GCN2, a protein kinase essential for translational derepression of *GCN4* mRNA in *Saccharomyces cerevisiae* [I. Roussou, G. Thireos, B. M. Hauge, *Mol. Cell. Biol.* **8**, 2132 (1988)].
10. C. Chothia and A. M. Lesk, *EMBO J.* **5**, 823 (1986).
11. R. K. Wierenga and W. G. J. Hol, *Nature* **302**, 842 (1983).
12. M. J. E. Sternberg and W. R. Taylor, *FEBS Lett.* **175**, 387 (1984).
13. M. P. Kamps and B. M. Sefton, *Mol. Cell. Biol.* **6**, 751 (1986).
14. M. J. Zoller, N. C. Nelson, S. S. Taylor, *J. Biol. Chem.* **256**, 10837 (1981).
15. M. P. Kamps, S. S. Taylor, B. M. Sefton, *Nature* **310**, 589 (1984).
16. M. W. Russo, T. J. Lukas, S. Cohen, J. V. Staros, *J. Biol. Chem.* **260**, 5205 (1985).
17. M. A. Snyder, J. M. Bishop, J. P. McGrath, A. D. Levinson, *Mol. Cell. Biol.* **5**, 1772 (1985).
18. M. Hannink and D. J. Donoghue, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 7894 (1985).
19. G. Weinmaster, M. J. Zoller, T. Pawson, *EMBO J.* **5**, 69 (1986).
20. W. S. Chen *et al.*, *Nature* **328**, 820 (1987); A. M. Honegger *et al.*, *Cell* **51**, 199 (1987).
21. C. K. Chou *et al.*, *J. Biol. Chem.* **262**, 1842 (1987).
22. L. T. Williams, personal communication.
23. S. Brenner, *Nature* **329**, 21 (1987).
24. D. Bryant and J. T. Parsons, *J. Virol.* **45**, 1211 (1983); *Mol. Cell. Biol.* **4**, 862 (1984).
25. H. N. Bramson *et al.*, *J. Biol. Chem.* **257**, 10575 (1982).
26. S. Shoji *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 848 (1981); S. Shoji, L. H. Ericsson, K. A. Walsh, E. H. Fischer, K. Titani, *Biochemistry* **22**, 3702 (1983).
27. G. Weinmaster, M. J. Zoller, M. Smith, E. Hinze, T. Pawson, *Cell* **37**, 559 (1984); G. Weinmaster and T. Pawson, *J. Biol. Chem.* **261**, 328 (1986); T. E. Kmiecik and D. Shalloway, *Cell* **49**, 65 (1987); H. Piwnica-Worms, K. B. Saunders, T. M. Roberts, A. E. Smith, S. H. Cheng, *ibid.*, p. 75; R. Herrera and O. M. Rosen, *J. Biol. Chem.* **261**, 11980 (1986); L. Ellis *et al.*, *Cell* **45**, 721 (1986).
28. R. A. Lindberg, unpublished data.
29. A. Lörincz and S. I. Reed, *Mol. Cell. Biol.* **6**, 4099 (1986).
30. J. M. Sowadski, N. h. Xuong, D. Anderson, S. S. Taylor, *J. Mol. Biol.* **182**, 617 (1984).
31. R. F. Doolittle, in *The Proteins*, H. Neurath and R. L. Hill, Eds. (Academic Press, New York, ed. 3, 1979), vol. 4, pp. 1–118.
32. D. Beach, B. Durkacz, P. Nurse, *Nature* **300**, 706 (1982); R. Booher and D. Beach, *Mol. Cell. Biol.* **6**, 3523 (1986).
33. M. G. Lee and P. Nurse, *Nature* **327**, 31 (1987).
34. L. Brizuela, G. Draetta, D. Beach, *EMBO J.* **6**, 3507 (1987).
35. M. D. Uhler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1300 (1986).
36. M. O. Showers and R. A. Maurer, *J. Biol. Chem.* **261**, 16288 (1986).
37. M. D. Uhler, J. C. Chrivia, G. S. McKnight, *ibid.*, p. 15360.
38. J. F. Cannon and K. Tatchell, *Mol. Cell. Biol.* **7**, 2653 (1987).
39. T. Toda, S. Cameron, P. Sass, M. Zoller, M. Wigler, *Cell* **50**, 277 (1987).
40. J. Lisziewicz, A. Godany, H.-H. Forster, H. Kuntzel, *J. Biol. Chem.* **262**, 2549 (1987).
41. K. Takio *et al.*, *Biochemistry* **23**, 4207 (1984).
42. P. J. Parker *et al.*, *Science* **233**, 853 (1986).
43. S. Ohno *et al.*, *Nature* **325**, 161 (1987).
44. L. Coussens *et al.*, *Science* **233**, 859 (1986).
45. J. L. Knopf *et al.*, *Cell* **46**, 491 (1986).
46. G. M. Housey, C. A. O'Brian, M. D. Johnson, P. Kirschmeier, I. B. Weinstein, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1065 (1987).
47. A. Rosenthal *et al.*, *EMBO J.* **6**, 433 (1987).
48. R. M. Hanley *et al.*, *Science* **237**, 293 (1987); C. R. Lin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5962 (1987).
49. M. K. Bennett and M. B. Kennedy, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 1794 (1987).
50. E. M. Reimann *et al.*, *Biochemistry* **23**, 4185 (1984); E. F. da Cruz e Silva and P. T. W. Cohen, *FEBS Lett.* **220**, 36 (1987).
51. J. S. Chamberlain, P. VanTuinen, A. A. Reeves, B. A. Philip, C. T. Caskey, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2886 (1987).
52. K. Takio *et al.*, *Biochemistry* **24**, 6028 (1985).
53. V. Guerriero, Jr., M. A. Russo, N. J. Olson, J. A. Putkey, A. R. Means, *ibid.* **25**, 8372 (1986).
54. The entire catalytic domain sequence has been determined by J. R. Woodgett (unpublished data).
55. J. L. Celenza and M. Carlson, *Science* **233**, 1175 (1986).

56. P. Russell and P. Nurse, *Cell* **49**, 569 (1987).
57. A. T. Lörincz and S. I. Reed, *Nature* **307**, 183 (1984).
58. J. Hindley and G. A. Phear, *Gene* **31**, 129 (1984).
59. The entire catalytic domain sequence has been determined by S. K. Hanks (unpublished data).
60. M. Simon, B. Seraphin, G. Faye, *EMBO J.* **5**, 2697 (1986).
61. K. Takio *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4851 (1987).
62. A. Saxena, R. Padmanabha, C. V. C. Glover, *Mol. Cell. Biol.* 7, 3409 (1987).
63. T. I. Bonner *et al.*, *Nucleic Acids Res.* **14**, 1009 (1986).
64. T. W. Beck, M. Huleihel, M. Gunnell, T. I. Bonner, U. R. Rapp, *ibid.* **15**, 595 (1987).
65. M. Huleihel *et al.*, *Mol. Cell. Biol.* **6**, 2655 (1986).
66. G. E. Mark, T. W. Seeley, T. B. Shows, J. D. Mountz, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6312 (1986).
67. R. Watson, M. Oskarsson, G. F. Vande Woude, *ibid.* **79**, 4078 (1982).
68. C. Van Beveren *et al.*, *Nature* **289**, 258 (1981).
69. F. A. Van der Hoorn and J. Firzlaff, *Nucleic Acids Res.* **12**, 2147 (1984).
70. M. A. Teague *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7371 (1986).
71. G. Boguslawski and J. O. Polazzi, *ibid.* **84**, 5848 (1987).
72. M. Patterson, R. A. Sclafani, W. L. Fangman, J. Rosamond, *Mol. Cell. Biol.* **6**, 1590 (1986).
73. P. Russell and P. Nurse, *Cell* **49**, 559 (1987).
74. M. McLeod and D. Beach, *EMBO J.* **5**, 3665 (1986).
75. G. Selten *et al.*, *Cell* **46**, 603 (1986).
76. D. J. McGeoch and A. J. Davison, *Nucleic Acids Res.* **14**, 1765 (1986).
77. S. K. Anderson, C. P. Gibbs, A. Tanaka, H.-J. Kung, D. J. Fujita, *Mol. Cell. Biol.* **5**, 1122 (1985); A. Tanaka *et al.*, *ibid.* 7, 1978 (1987).
78. R. Martinez, B. Mathey-Prevot, A. Bernards, D. Baltimore, *Science* **237**, 411 (1987).
79. T. Takeya and H. Hanafusa, *Cell* **32**, 881 (1983).
80. R. E. Steele, *Nucleic Acids Res.* **13**, 1747 (1985).
81. J. Sukegawa *et al.*, *Mol. Cell. Biol.* 7, 41 (1987).
82. M. Nishizawa *et al.*, *ibid.* **6**, 511 (1986); R. C. Parker, G. Mardon, R. V. Lebo, H. E. Varmus, J. M. Bishop, *ibid.* **5**, 831 (1985).
83. K. Inoue *et al.*, *Oncogene* **1**, 301 (1987).
84. K. Semba *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5459 (1986); T. Kawakami, C. Y. Pennington, K. C. Robbins, *Mol. Cell. Biol.* **6**, 4195 (1986).
85. Y. Yamanashi *et al.*, *Mol. Cell. Biol.* 7, 237 (1987).
86. J. M. Trevillyan *et al.*, *Biochim. Biophys. Acta* **888**, 286 (1986); Y. Koga *et al.*, *Eur. J. Immunol.* **16**, 1643 (1986).
87. J. D. Marth, R. Peet, E. G. Krebs, R. M. Perlmutter, *Cell* **43**, 393 (1985); A. F. Voronova and B. M. Sefton, *Nature* **319**, 682 (1986).
88. N. Quintrell *et al.*, *Mol. Cell. Biol.* 7, 2267 (1986); S. F. Ziegler, J. D. Marth, D. B. Lewis, R. M. Perlmutter, *ibid.*, p. 2276.
89. M. A. Simon, B. Drees, T. Kornberg, J. M. Bishop, *Cell* **42**, 831 (1985).
90. F. M. Hoffman, L. D. Fresco, H. Hoffman-Falk, B.-Z. Shilo, *ibid.* **35**, 393 (1983).
91. R. J. Gregory, K. L. Kammermeyer, W. S. Vincent III, S. G. Wadsworth, *Mol. Cell. Biol.* 7, 2119 (1987).
92. E. Shtivelman, B. Lifshitz, R. P. Gale, B. A. Roe, E. Canaani, *Cell* **47**, 277 (1986).
93. G. D. Kruh *et al.*, *Science* **234**, 1545 (1986).
94. J. M. Goddard, J. J. Weiland, M. R. Capecchi, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 2172 (1986).
95. A. J. M. Roebroek *et al.*, *EMBO J.* **4**, 2897 (1985).
96. A. J. M. Roebroek, J. A. Schalken, C. Onnekink, H. P. J. Bloemers, W. J. M. Van de Ven, *J. Virol.* **61**, 2009 (1987).
97. C.-C. Huang, C. Hammond, J. M. Bishop, *J. Mol. Biol.* **181**, 175 (1985).
98. A. Ullrich *et al.*, *Nature* **309**, 418 (1984).
99. L. Coussens *et al.*, *Science* **230**, 1132 (1985); T. Yamamoto *et al.*, *Nature* **319**, 230 (1986).
100. C. I. Bargmann, M.-C. Hung, R. A. Weinberg, *Nature* **319**, 226 (1986).
101. E. Livneh, L. Glazer, D. Segal, J. Schlessinger, B.-Z. Shilo, *Cell* **40**, 599 (1985).
102. A. Ullrich *et al.*, *Nature* **313**, 756 (1985); Y. Ebina *et al.*, *Cell* **40**, 747 (1985).
103. A. Ullrich *et al.*, *EMBO J.* **5**, 2503 (1986).
104. Y. Nishida, M. Hata, Y. Nishizuka, W. J. Rutter, Y. Ebina, *Biochem. Biophys. Res. Commun.* **141**, 474 (1986).
105. H. Matsushime, L.-H. Wang, M. Shibuya, *Mol. Cell. Biol.* **6**, 3000 (1986); C. Birchmeier, D. Birnbaum, G. Waitches, O. Fasano, M. Wigler, *ibid.*, p. 3109.
106. W. S. Neckameyer, M. Shibuya, M.-T. Hsu, L.-H. Wang, *ibid.*, p. 1478.
107. S. B. Podell and B. M. Sefton, *Oncogene* **2**, 9 (1987).
108. E. Hafen, K. Basler, J.-E. Edstroem, G. M. Rubin, *Science* **236**, 55 (1987).
109. D. Martin-Zanca, S. H. Hughes, M. Barbacid, *Nature* **319**, 743 (1986).
110. M. Park *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6379 (1987); A. M.-L. Chan *et al.*, *Oncogene* **1**, 229 (1987).
111. Y. Yarden *et al.*, *Nature* **323**, 226 (1986).
112. L. Coussens *et al.*, *ibid.* **320**, 277 (1986).
113. Y. Yarden *et al.*, *EMBO J.* **6**, 3341 (1987).
114. M. Takahashi and G. M. Cooper, *Mol. Cell. Biol.* 7, 1378 (1987).
115. D. A. Foster, J. B. Levy, G. Q. Daley, M. C. Simon, H. Hanafusa, *ibid.* **6**, 325 (1986).
116. D.-F. Feng, M. S. Johnson, R. F. Doolittle, *J. Mol. Evol.* **21**, 112 (1985).
117. W. M. Fitch and E. Margoliash, *Science* **155**, 279 (1967).
118. D.-F. Feng and R. F. Doolittle, *J. Mol. Evol.* **25**, 351 (1987).

119. We thank D.-F. Feng and R. F. Doolittle for much helpful advice concerning phylogenetic tree construction, R. W. Holley for support, J. R. Woodgett for making unpublished sequence data available, R. A. Lindberg and L. T. Williams for allowing us to reference their unpublished work, K. Hyde for assistance in sequence data entry, and L. Norris for help in preparation of Fig. 1. Supported by grant GM38793 from the NIH (SKH).